

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: June 24, 2003, 19:13:28 ; Search time 62 Seconds

(without alignments)
5721.180 Million cell updates/sec

Title: US-09-830-244b-2

Sequence: 1 ccatgtttttttagatcaaaag.....gccttttccatgctactg 1331

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 908470 segs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+_n2p_model -DEV=xlp
-O=/cgg2_1/USPRO.spool/US09830244.funat.24062003.130000.10004/app_query.fasta.1.1479
-DB=geneseq.101002 -OEMT=fastcan -SUFFIX=n2p_rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blonsum62 -TRANS=human40.cdl
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MAINLEN=0 -MAXLEN=2000000000
-USER=US09830244.GCCH.1.1.114 -funat.24062003.130000.10004 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -MARN_TIMECUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: A.Geneseq.101002.*
- 2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 21: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 24: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	522	22.5	95	21	AA94526	Human lysine-rich
2	136	5.9	62	21	AA94527	Human statherin pr
3	136	5.9	82	21	AA94528	Human statherin pr
4	93	4.0	384	20	AA94529	Human statherin pr
5	93	4.0	384	20	AA94530	Human statherin pr
6	93	4.0	384	20	AA94531	Human statherin pr
7	93	4.0	384	20	AA94532	Human statherin pr
8	87.5	3.7	360	16	AA94533	Human statherin pr
9	85.5	3.7	360	16	AA94534	Human statherin pr
10	85.5	3.7	360	16	AA94535	Human statherin pr
11	85.5	3.7	360	16	AA94536	Human statherin pr
12	85.5	3.7	360	16	AA94537	Human statherin pr
13	85.5	3.7	360	16	AA94538	Human statherin pr
14	85.5	3.7	360	16	AA94539	Human statherin pr
15	85.5	3.7	360	16	AA94540	Human statherin pr
16	85.5	3.7	360	16	AA94541	Human statherin pr
17	85.5	3.7	360	16	AA94542	Human statherin pr
18	85.5	3.7	360	16	AA94543	Human statherin pr
19	85.5	3.7	360	16	AA94544	Human statherin pr
20	85.5	3.7	360	16	AA94545	Human statherin pr
21	85.5	3.7	360	16	AA94546	Human statherin pr
22	85.5	3.7	360	16	AA94547	Human statherin pr
23	85.5	3.7	360	16	AA94548	Human statherin pr
24	85.5	3.7	360	16	AA94549	Human statherin pr
25	85.5	3.7	360	16	AA94550	Human statherin pr
26	85.5	3.7	360	16	AA94551	Human statherin pr
27	85.5	3.7	360	16	AA94552	Human statherin pr
28	85.5	3.7	360	16	AA94553	Human statherin pr
29	85.5	3.7	360	16	AA94554	Human statherin pr
30	85.5	3.7	360	16	AA94555	Human statherin pr
31	85.5	3.7	360	16	AA94556	Human statherin pr
32	85.5	3.7	360	16	AA94557	Human statherin pr
33	85.5	3.7	360	16	AA94558	Human statherin pr
34	85.5	3.7	360	16	AA94559	Human statherin pr
35	85.5	3.7	360	16	AA94560	Human statherin pr
36	85.5	3.7	360	16	AA94561	Human statherin pr
37	85.5	3.7	360	16	AA94562	Human statherin pr
38	85.5	3.7	360	16	AA94563	Human statherin pr
39	85.5	3.7	360	16	AA94564	Human statherin pr
40	85.5	3.7	360	16	AA94565	Human statherin pr
41	85.5	3.7	360	16	AA94566	Human statherin pr
42	85.5	3.7	360	16	AA94567	Human statherin pr
43	85.5	3.7	360	16	AA94568	Human statherin pr
44	85.5	3.7	360	16	AA94569	Human statherin pr
45	85.5	3.7	360	16	AA94570	Human statherin pr

ALIGNMENTS

RESULT 1
ID AA94526 standard: protein; 95 AA.
AA94526;
06-OCT-2000 (first entry)
Human lysine-rich statherin protein.
Lysine-rich statherin protein; LSP; acidic phosphoprotein; human;
precipitation inhibitor; autoimmune; inflammatory disorder; AIDS;
asthma; allergy; diabetes mellitus; fungal; bacterial infection;
cancer; leukemia; adenocarcinoma; melanoma.
Homo sapiens.
Key Modified-site 15 Location/Qualifiers

FT /note- "Potential casein kinase II phosphorylation
FT site"
FT Modified-site 24
FT /note- "Potential N-glycosylation site"
FT Modified-site 29
FT /note- "Potential protein kinase C phosphorylation
FT site"
FT Modified-site 76
FT /note- "Potential N-glycosylation site"
XX MO200024779-A1.
XX 04-MAY-2000.
XX 22-OCT-1999; 99MO-US24046.
XX 23-OCT-1998; 98US-0155209.
XX (INCY-) INCYTE PHARM INC.
XX Tang YT, Corley NC, Guegler KJ, Patterson C;
XX MPI: 2000-350699/30.
XX N-PSDB; AAA48963.
XX Purified polypeptide used for treating or preventing a disorder
XX characterized by expression or activity of lysine-rich statherin
XX proteins -
XX
XX Claim 1; Page 69; 75pp; English.
XX
XX The present sequence is human lysine-rich statherin protein (LRSP). The
XX cDNA sequence encoding this protein was identified through analysis of
XX a cDNA library of breast tumour tissue (BRSTN014). The LRSP sequence
XX was found to have homology with human statherin (AAV94527) and human
XX basic histidine-rich protein (AAV94528). Human statherin is a
XX phosphoprotein that acts as an inhibitor of precipitation of calcium
XX phosphate salts in the oral cavity. The LRSP polypeptide and its
XX antagonists may be useful for treating or preventing disorders
XX associated with the activity of LRSP. Such disorders include
XX autoimmune/inflammatory disorders (for example AIDS, allergies, asthma,
XX diabetes mellitus), bacterial and fungal infection and cancers (such
XX as leukemia, adenocarcinoma, melanoma). Antibodies to LRSP may be
XX useful for diagnosis of the above disorders.
XX
XX Sequence 95 AA:
SQ
Alignment Scores:
Pred. No.: 6,75e-56 Length: 95
Score: 522.00 Matches: 95
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 22.48% Indels: 0
Gaps: 0
US-09-830-244B-2 (1-1331) x AAV94526 (1-95)
OY 871 ATGTGCTTCTAATAAGTGGGAGAAACAGCATTTTAAGTAAGTCTTTGGAGACATGAT 930
DB 1 MetTyrPheHisLysValIGlyArgLysGlnHisPheLysValThrPheTrpGluThrAsp 20
OY 931 TTGAGTAAATATAAATCTGGTCTGCTTAAAGAAAAAAGCCCTTCCACCTTACTGT 990
DB 21 LeuSerLysHisLysThrLeuValSerLeuLysLysLysProPheHisLeuTyrCys 40
OY 991 GTCATTTAATCCCTTACTGTCACAAAGTTAATATCTTATTTGGATTTGCTTTATA 1050
DB 41 ValIleTyrLleProLeuValProLysLeuIleLeuPheLeuLysPheIleAlaPheIle 60
OY 1051 CCAAGAGCCTTTTCAGCCAGTTCAGACAGACACATATACGACACACATACACACAC 1110
DB 61 ProLysSerLeuLysSerGlnPheGlnAsnHisLysTyrThrHisAsnHisThrAsnHis 80

OY 1111 AATACCAACAATATACGTTTAAATATCATCATGACTACTGACAGACA 1155
DB 81 AsnThrAsnAsnIleArgPheAsnIleIleSerAsnCysArgThr 95
RESULT 2
ID AAV94527
AAV94527 standard; protein; 62 AA.
AC AAV94527;
XX 06-OCT-2000 (first entry)
XX Human statherin protein.
XX
XX Lysine-rich statherin protein; LRSP; acidic phosphoprotein; human;
XX precipitation inhibitor; autoimmune; inflammatory disorder; AIDS;
XX asthma; allergy; diabetes mellitus; fungal; bacterial infection;
XX cancer; leukemia; adenocarcinoma; melanoma.
XX Homo sapiens.
XX
XX MO200024779-A1.
XX 04-MAY-2000.
XX 22-OCT-1999; 99MO-US24046.
XX 23-OCT-1998; 98US-0155209.
XX (INCY-) INCYTE PHARM INC.
XX
XX Tang YT, Corley NC, Guegler KJ, Patterson C;
XX MPI: 2000-350699/30.
XX N-PSDB; AAA48964.
XX Purified polypeptide used for treating or preventing a disorder
XX characterized by expression or activity of lysine-rich statherin
XX proteins -
XX
XX Disclosure; Page 70; 75pp; English.
XX
XX The present invention relates to human lysine-rich statherin protein
XX (LRSP) (AAV94526). The cDNA sequence encoding this protein was identified
XX through analysis of a cDNA library of breast tumour tissue (BRSTN014).
XX The LRSP sequence was found to have homology with human statherin (the
XX present sequence) and human basic histidine-rich protein (AAV94528).
XX Human statherin is a phosphoprotein that acts as an inhibitor of
XX precipitation of calcium phosphate salts in the oral cavity. The LRSP
XX polypeptide and its antagonists may be useful for treating or preventing
XX disorders associated with the activity of LRSP. Such disorders include
XX autoimmune/inflammatory disorders (for example AIDS, allergies, asthma,
XX diabetes mellitus), bacterial and fungal infection and cancers (such as
XX leukemia, adenocarcinoma, melanoma). Antibodies to LRSP may be useful
XX for diagnosis of the above disorders.
XX
XX Sequence 62 AA:
SQ
Alignment Scores:
Pred. No.: 8.24e-08 Length: 62
Score: 136.00 Matches: 24
Percent Similarity: 85.71% Conservative: 0
Best Local Similarity: 85.71% Mismatches: 4
Query Match: 5.86% Indels: 0
Gaps: 0
US-09-830-244B-2 (1-1331) x AAV94527 (1-62)
OY 1047 TATACCAAGAGCCTTATATAGCCAGTTCAGACAAACATATACGACACACATACCA 1106
DB 35 TyrGlyTyrGlyProTyrGlnProValProGluGlnProLeuTyrProGlnProTyrGln 54
OY 1107 CCACATACCAACAATATACGTTT 1130

Db	55	ProGlnTyrGlnGlnTyrThrPhe 62
RESULT 3		
ID	AA042456	
XX	AA042456 standard; Protein; 82 AA.	
AC	AA042456;	
XX	08-FEB-2001 (first entry)	
XX		
DE	Human ORFX ORF220 polypeptide sequence SEQ ID NO:4440.	
KW	Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;	
KW	vulnerary; antiposrotatic; antiparkinsonian; neurotropic; neuroprotective;	
KW	immunostimulant; osteopathic; antiarthritic; immunosuppressant; cardiac;	
KW	hypotensive; dermatolytic; coagulant; vasotropic; antidiabetic;	
KW	antiviral; antibacterial; immunosuppressive; antineoplastic;	
KW	antitumoric; gene therapy; cancer; proliferative disorder; hypotension;	
KW	neurodegenerative disorder; osteoarthritis; graft vs host disease;	
KW	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;	
KW	cholesterol ester storage; systemic lupus erythematosus; infection;	
KW	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;	
KW	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;	
KW	bone damage; cartilage damage; antineoplastic disease; coagulation;	
KW	thrombosis; contraceptive.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200058473-A2.	
XX		
PD	05-OCT-2000.	
XX		
PF	31-MAR-2000; 2000WO-US08621.	
PR	31-MAR-1999; 99US-0127607.	
PR	02-APR-1999; 99US-0127636.	
PR	05-APR-1999; 99US-0127728.	
PR	30-MAR-2000; 2000US-0540763.	
XX		
PA	(CUBA-) CRAGEN CORP.	
XX		
PI	Shimkets RA, Leach M;	
DR	WPI; 2000-602362/57.	
DR	N-PSDB; AAC76665.	
XX		
PT	Novel nucleic acids and peptides derived from open reading frame X,	
PT	useful for treating e.g. cancers, proliferative disorders,	
PT	neurodegenerative disorders and cardiovascular disease -	
XX		
PS	Claim 11; Page 3632; 5507pp; English.	
XX		
CC	AAC74446 to AAC7606 encode the proteins given in AAB40237 to AAB43397,	
CC	which represent the human ORFX open reading frames 1 to 3161. The ORFX	
CC	sequences have activities such as: cytostatic; hepatotropic; vulnerary;	
CC	antiposrotatic; antiparkinsonian; neurotropic; neuroprotective;	
CC	osteopathic; anticonvulsant; antiarthritic; immunosuppressant;	
CC	immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;	
CC	antidiabetic; hypotensive; dermatolytic; immunosuppressive;	
CC	antitumoric; antibacterial; antiviral; antifungal; antineoplastic;	
CC	antithyroid; and antitumoric. The sequences can be used for determining	
CC	the presence of or predisposition to, or preventing or treating	
CC	pathological conditions associated with an ORFX-associated disorder. The	
CC	nucleic acids can be used to express ORFX proteins in gene therapy	
CC	vectors. The proteins and nucleic acids may be used to treat cancers,	
CC	proliferative disorders, neurodegenerative disorders, diabetes mellitus,	
CC	graft vs host disease, cardiovascular disease, diabetes mellitus,	
CC	hypertension, hypothyroidism, cholesterol ester storage, systemic lupus	
CC	erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,	
CC	bacterial or fungal infection, malaria, autoimmune disorders, asthma,	
CC	allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,	

CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX

Sequence 82 AA;

Alignment Scores:

Pred. No.:	9,26e-08	Length:	82
Score:	136.00	Matches:	24
Percent Similarity:	85.71%	Conservative:	0
Best Local Similarity:	85.71%	Mismatches:	4
Query Match:	5.86%	Indels:	0
DB:	21	Gaps:	0

US-09-830-244B-2 (1-1331) x AAB42456 (1-82)

QY 1047 TATACCAAGAGCCTTATCAGCGACTTCAGACACACCATATACGACCAATACCCA 1106
||| |||||||
Db 55 TTYGlyTYrGlyProTyrGlnProValProGlnGlnProLeuTyrProGlnProTyrCln 74
|||||
QY 1107 CCACATATACCAACATATACGTTT 1130
|||||
Db 75 ProGlnTyrGlnGlnTyr-ThrPhe 82
|||||

RESULT 4
AAAY00147
ID AAY00147 standard; Protein; 384 AA.
XX
XX AAY00147;
AC
XX
XX
DT 20-APR-1999 (first entry)
XX
XX
DE Enterococcus faecalis antigenic polypeptide fragment EF072.
XX
XX Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
XX detection; attenuation; antigenic.
XX
XX Enterococcus faecalis.
OS
XX
PN W09850554-A2.
PD 12-NOV-1998.
XX
XX
PF 04-MAY-1998; 98WO-US08959.
XX
XX
PR 14-NOV-1997; 97US-0066009.
PR 06-MAY-1997; 97US-0044031.
PR 16-MAY-1997; 97US-0046655.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PI
XX
XX Bailey C, Choi GH, Hromockyj A, Kunsch CA;
DR WPI; 1999-070095/06.
XX
XX N-PSDB; AAX20137.
PS
PS
XX
XX

Claim 9; Page 163; 301pp; English.

PT The present sequence represents an antigenic polypeptide fragment
PT isolated from *Enterococcus faecalis*. The present invention describes
PT genes, proteins and antigenic polypeptides isolated from *E. faecalis*.
PT The proteins can be used in vaccines for preventing or attenuating an
PT infection caused by a member of the *Enterococcus* genus in an animal.
PT They can also be used for detecting *Enterococcus* antibodies in a sample.
PT The nucleotide sequences can be used for detecting *Enterococcus* nucleic
PT acids. Products from the present invention can also be used for
PT screening compounds to identify agonists and antagonists of *E. faecalis*
PT protein activity.
XX
XX
XX

Sequence 384 AA;

Alignment Scores:

Pred. No.:	0.0409	Length:	384
Score:	93.00	Matches:	74
Percent Similarity:	36.26%	Conservative:	21
Best Local Similarity:	28.24%	Mismatches:	85
Query Match:	4.02%	Indels:	83
DB:	20	Gaps:	12

US-09-830-244B-2 (1-1331) x AAY00147 (1-384)

```

QY 672 ACTGTCCTGCTCCAGAGGGATTCATGAGATGCTAAGAGGAAACACCTGCAG 613
DB 149 ThrValLeuAspLysGlnGlnGlyPheasn-----Gln 159
QY 612 GGAAGCAGCATCTGATTCGATTAAG-----TAC 580.
DB 160 Gly---GlnHisIleasnTrpGlnLeuThrGlnIleProAlaAsnIleuGlyTyr 178
QY 579 CAAAAAATATTAAGCTATCTCTTCGACACATGCGCTTCCTGCTTAACACTTCAGT 520
DB 179 Gln-GluPheArgLeuSerAspLysAlaAspThrLeuThrLeuProGluSerIle 198
QY 519 CATTCAAAAGCCTATTGCTAGAG-----TCAGCAAAATCCAGTG 479
DB 198 egluValIysValAlaGlyLysThrValThrThrGlyTyrThrLeuThrGlnLysHis 218
QY 478 TAGATTACTGTACAGATTATTATTA-----ACTATGCT 446
DB 218 sglYpHeThrLeuAspPheSerIleLysAspLeuGlnAsnPhelaAsnGlnThrMetH 238
QY 445 GTATTCA-----CCAATAAGCGCATTCACCAATAATAGCA----- 412
DB 238 rValSerTyrGlnMetArgLeuGlnLysThrAlaGluProAspThrAlaIleAsnGln 258
QY 411 -----TTAGGCATCAAAAGTCGAAGATCTGAATGCGACATTCGCCACAAAG 359
DB 258 uglGlnLeuValThrAspLysHisThrLeuThrLysArgAlaThrValArgThrGlyG 278
QY 358 AAG-----AAATCCATGCTCCTATA----- 336
DB 278 YLysSerPheValIysValAspSerGluAsn-AlaLysIleThrLeuProGluAlaValP 298
QY 335 -----AAGGAGATATATGCAATGCGACAGCAAGTAAATTTCTGTC 293
DB 298 heIleValIysAsnGlnAlaGlyLysThrLeuAsnGlnThrAlaAsnGlyTyrArgTrp 318
QY 292 AAAAGAAAGTACTTCTCAAGCAATCAGAAACCTGTCCCAATTAAGCGCTGGAGC 233
DB 318 InLysGlu-----LysAlaLeuAlaLysLysPheThrSerAsnGlnAlaGlyGlu 334
QY 232 CACATTAGCCACTATCTCAGATACTGATATTGTTTACCTTACGGGCGCTCAATATTC 173
DB 335 -----PheSerValLysGly**LysArgT 343
QY 172 CCCCTCATCTGCTAAACTCACTGTTTACTCACAAGTGTGATTAATTCACGCAAGTA 113
DB 343 rPrroValLeuLeuGlnGlyArgAsnLeuGlyThrLysArgLeuSer-----SerGlnSerA 361
QY 112 ACCTTAACATTCCTCTATGCGAGTGCAGAAATTCAGACATTTGTAACACTGTAATTTCA 53
DB 361 snArgAsnSerPheTyrGlyGlyLysPheLeuGlyAsnGlnLysThrThrAsnSer 381
QY 52 CG 51
DB 381 hr 381

```

RESULT 5
ABP43366 standard; Protein; 384 AA.

XX AC ABP43366;
XX

DT 05-AUG-2002 (first entry)

DE E faecalis EF072 antigenic fragment.

XX Enterococcus: vaccine; gastrointestinal disease; diagnosis; antibiotic.

XX Enterococcus faecalis.

XX US2002045737-A1.

XX 18-APR-2002.

XX 04-MAY-1998; 98US-0071035.

XX 04-MAY-1998; 98US-0071035.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Chol GH, Bailey C, Hromockyj A, Kunsch CA;

XX WPI; 2002-425450/45.

XX N-PSDB; ABW98122.

PT New genes and polypeptides from Enterococcus faecalis, useful as
 PT vaccines for preventing, treating or attenuating an infection caused by
 PT a member of the Enterococcus genus in an animal, particularly E. faecalis
 PT faecalis

PS Claim 9; Page 133; 255pp; English.

CC The present invention provides the protein and coding sequences of a
 CC number of polypeptides from Enterococcus faecalis. The proteins can be
 CC used as vaccines for preventing or attenuating an infection caused by a
 CC member of the Enterococcus genus in an animal, particularly E. faecalis.
 CC The polynucleotide is also useful for preventing or treating E. faecalis
 CC infection. The present sequence is a protein of the invention.

SQ Sequence 384 AA;

Alignment Scores:

Pred. No.:	0.0409	Length:	384
Score:	93.00	Matches:	74
Percent Similarity:	36.26%	Conservative:	21
Best Local Similarity:	28.24%	Mismatches:	85
Query Match:	4.02%	Indels:	83
DB:	23	Gaps:	12

US-09-830-244B-2 (1-1331) x ABP43366 (1-384)

```

QY 672 ACTGTCCTGCTCCAGAGGGATTCATGAGATGCTAAGAGGAAACACCTGCAG 613
DB 149 ThrValLeuAspLysGlnGlnGlyPheasn-----Gln 159
QY 612 GGAAGCAGCATCTGATTCGATTAAG-----TAC 580
DB 160 Gly---GlnHisIleasnTrpGlnLeuThrGlnIleProAlaAsnIleuGlyTyr 178
QY 579 CAAAAAATATTAAGCTATCTCTTCGACACATGCGCTTCCTGCTTAACACTTCAGT 520
DB 179 Gln-GluPheArgLeuSerAspLysAlaAspThrLeuThrLeuProGluSerIle 198
QY 519 CATTCAAAAGCCTATTGCTAGAG-----TCAGCAAAATCCAGTG 479
DB 198 egluValIysValAlaGlyLysThrValThrThrGlyTyrThrLeuThrGlnLysHis 218
QY 478 TAGATTACTGTACAGATTATTATTA-----ACTATGCT 446
DB 218 sglYpHeThrLeuAspPheSerIleLysAspLeuGlnAsnPhelaAsnGlnThrMetH 238
QY 445 GTATTCA-----CCAATAAGCGCATTCACCAATAATAGCA----- 412
DB 238 rValSerTyrGlnMetArgLeuGlnLysThrAlaGluProAspThrAlaIleAsnGln 258

```



```

QY 411 -----TTAGTCACATCAAAAGATCTGAATGCTAATGGACATTTCCACAAGG 359
Db 258 uclglnleuValThrAspLysIsthrLeuThrLysArgAlaThrValArgThrGlyGI 278
QY 358 AAAG-----AAATCCATGGCTCCTATA----- 336
Db 278 yLysSerPheValLysValAspSerGluAsn-AlaLysIleThrLeuProGluAlaValP 298
QY 335 -----AAGGAGATATATAGCAATGGCAACCAAGATGAATATTTCTGTC 293
Db 298 heileValLysAsnGlnAlaGlyGluTyrLeuAsnGluThrAlaAsnGlyTyrArgTrpG 318
QY 292 AAAAGAAAGTACTTCTCAACCAATCAGAAAAAATGTTCCCAATTAATGAGCCGAGGC 233
Db 318 InLysGlu-----LysAlaLeuAlaLysLysPheThrSerAsnGlnAlaGlyGlu- 334
QY 232 CACATTAGCCACATCTCAGATGACTGATTTATTTGTTAGCCTTAGGGGCTCAATATTC 173
Db 335 -----PheSerValLysGly**LysArgT 343
QY 172 CCCCTCATCTGCTAAACTCAACTGTTTACTCAGACATGATTAATTCACGCAAGTA 113
Db 343 rPProValLeuLeuGlyArgAsnLeuLysThrLysArgLeuSer-----SerGluSerA 361
QY 112 ACCTTAACATTCCTCATGCGCAGTGCAGAAATTCAGACATTTGTAACACTGTAATTTCA 53
Db 361 snArgAsnSerPheThrGlyGlyLysLysPheLeuLysAsnGluArgThrThrAsnSerT 381
QY 52 CG 51
Db 381 hr 381

```

RESULT 6

AAV00146 standard; Protein; 430 AA.

```

ID AAV00146 standard; Protein; 430 AA.
AC AAV00146;
XX 20-APR-1999 (first entry)
DE Enterococcus faecalis protein EF072.
XX Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
KM detection; attenuation; antigenic.
XX Enterococcus faecalis.
OS
XX MO9850554-A2.
PN 12-NOV-1998.
XX 04-MAY-1998; 98WO-US08959.
XX 14-NOV-1997; 97US-0066009.
XX 06-MAY-1997; 97US-0044031.
XX 16-MAY-1997; 97US-0046555.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Bailey C, Choi GH, Hromocky A, Kunsch CA;
XX WPI; 1999-070095/06.
XX N-PSDB: AAX20136.
XX New isolated Enterococcus faecalis polynucleotides - used to develop
XX products for the detection of Enterococcus and for use in vaccines
XX for prevention or attenuation of Enterococcus infection
XX Claim 9; Page 163; 301pp; English.
XX
XX The present sequence represents a protein isolated from
XX Enterococcus faecalis. The present invention describes genes, proteins
XX and antigenic polypeptides isolated from E. faecalis. The proteins can

```

CC be used in vaccines for preventing or attenuating an infection caused by a member of the Enterococcus genus in an animal. They can also be used for detecting Enterococcus antibodies in a sample. The nucleotide sequences can be used for detecting Enterococcus nucleic acids. CC products from the present for detecting Enterococcus nucleic acids. CC compounds to identify agonists and antagonists of E. faecalis protein activity.

US-09-830-244b-2 (1-1331) x AAV00146 (1-430)

QY 672 ACTGCTCTGCTCCAGAGGGATTCATGAGATCGTAAGAAACAACCTGACG 613
 Db 193 ThrValLeuAspLysGlnGlyPheAsn-----Gln 203

QY 612 GGGAGACAGACCTTGATTCAGATAAG-----TAC 580
 Db 204 Gly---GluHisIleAsnTyrGlnLeuThrThrGlnIleProAlaAsnIleLeuGlyTyr 222

QY 579 CAAATAATATTAAGCTATCTCTTCTGACACATGCTCTTCTGTAACCTTCACT 520
 Db 223 Gln-GluPheArgLeuSerAspLysAlaAspThrThrLeuThrLeuLeuProGluSerI 242

QY 519 CATTCAAAAGGCTATGCTAGACAG-----TCACGCAATTCACAGTG 479
 Db 242 eGluValLysValAlaGlyLysThrValThrThrGlyTyrThrLeuThrThrGlnLysH 262

QY 478 TAGATTCTGTAACAGTTTATTATTA-----ACTATGCT 446
 Db 262 sGlyPheThrLeuAspPheSerIleLysAspLeuGlnAsnPheAlaAsnGlnThrMeth 282

QY 445 GTATTC-----CAATTAACGGCATTTACCAATATGCA----- 412
 Db 282 rValSerTyrGlnMetArgLeuGluLysThrAlaGluProAspThrAlaIleAsnAsnG 302

QY 411 -----TTAGTCACATCAAAAGATCTGAATAATGCTAATGACATTTCCACAAAGG 359
 Db 302 uclglnleuValThrAspLysIsthrLeuThrLysArgAlaThrValArgThrGlyGI 322

QY 358 AAAG-----AAATCCATGGCTCCTATA----- 336
 Db 322 yLysSerPheValLysValAspSerGluAsn-AlaLysIleThrLeuProGluAlaValP 342

QY 335 -----AAGGAGATATATAGCAATGGCAACCAAGATGAATATTTCTGTC 293
 Db 342 heileValLysAsnGlnAlaGlyGluTyrLeuAsnGluThrAlaAsnGlyTyrArgTrpG 362

QY 292 AAAAGAAAGTACTTCTCAACCAATCAGAAAAAATGTTCCCAATTAATGAGCCGAGGC 233
 Db 362 InLysGlu-----LysAlaLeuAlaLysLysPheThrSerAsnGlnAlaGlyGlu- 378

QY 232 CACATTAGCCACATCTCAGATGACTGATTTATTTGTTAGCCTTAGGGGCTCAATATTC 173
 Db 379 -----PheSerValLysGly**LysArgT 387

QY 172 CCCCTCATCTGCTAAACTCAACTGTTTACTCAGACATGATTAATTCACGCAAGTA 113
 Db 387 rPProValLeuLeuGlyArgAsnLeuLysThrLysArgLeuSer-----SerGluSerA 405

QY 112 ACCTTAACATTCCTCATGCGCAGTGCAGAAATTCAGACATTTGTAACACTGTAATTTCA 53
 Db 405 snArgAsnSerPheThrGlyGlyLysLysPheLeuLysAsnGluArgThrThrAsnSerT 425

QY 52 CG 51

PT Identifying modulators of receptor activity using orthologues of human
PT receptors by contacting a compound with receptors from two species, and
XX measuring the effect of the compound on the receptors -
PS
XX Disclosure; Page 88-89; 108pp; English.

The invention relates to identifying modulators of receptor activity using orthologues of human receptors by contacting a compound with receptors from two species, and measuring the effect of the compound on the receptors. Included is an isolated nucleic acid (I) comprising a nucleotide sequence encoding bradykinin B1 receptor (II); or encoding a CXCR2 receptor (III). The method is useful for identifying a compound having dual specificity to modulate the activity of a desired polypeptide in two different species. The method is useful for identifying a non-human animal model for testing compounds with potential efficacy as human receptor modulators. The method comprises contacting a non-receptors from at least two species; measuring an effect of the compound on the receptors; and selecting an animal model representing a species having a receptor that exhibits the desired effect when contacted with the test compound. Also described is a method for identifying a compound that modulates (II) or (III) activity, where a compound is an agonist, antagonist, suppressor, inhibitor and inducer, useful in treatment of pain and inflammation and other receptor related pathologies. AA080476-AA080492 represent bradykinin I and CXCR2 receptor amino acid sequences of the invention.

Sequence 355 AA;

Alignment Scores:	
Pred. No.:	0, 192
Score:	87.50
Percent Similarity:	37.84%
Best Local Similarity:	22.76%
Query Match:	3.78%
:	23
:	35
Length:	355
Matches:	59
Conservative:	39
Mismatch:	87
Indels:	75
Gaps:	10

2445-2 (1-1331) X AA080490 (1-355)

[illegible]

QY	675	CTAAGTCCTGCGTCCAGAGGGCATTTCAATGAGAT-----	637
Dd	206	LeuArgIleLeuProGlnSerPheGlyPheIleValPProLeuLlleMeLeuPhCys	225
QY	636	-----CGTAAGAAGAAACAACCTGCAGGGGAACGACATCTTGATT	592
Dd	226	TyrGlyPheThrLeuArgThrLeuPheLysAlaHisMetGlyGlnLysAsnArgAlaMet	245
QY	591	CAGATAAGTACCAAAANAATTTAGCTATCTCTTTGCGACACATGCCCTTCTGTGTT	532
Dd	246	ArgValIlePheAlaValValLeuIlePheLeuLeucysTrp-LeuProLysSerLeuVa	265
QY	531	AACACTTCA-----GTATTCAAAAGCGTATTGGTAGA	498
Dd	265	IleuLeuAlaspthrLeuMetArgrthrGlnValIleGlnGluThnCysGlnArg	283

D1 04-APR-1993 (first entry)
XX

receptor polypeptide in 4AB.

XX **polypeptide; G-protein-coupled receptor**

Homo sapiens

PN WO9218641-A
XY

29-OCT-1992

10-APR-1992; 92WO-US02977.

10-APR-1991; 91US-0685101.

09-DEC-1991; 91US-0803842.

REPLIGEN CORP.
(NYSE: RGEN)
10000
10000

Navarro T

WPT: 1993-383133 44c

N-PSDB; AAQ30012.

interleukin-8 binding site

Disclosure: Eln 2: 7122: Eln

Rabbit hiah affinity 11-8

produce a rabbit neutrophil and u:

oligonucleotide (AA030015)

sequence derived from the second coupled receptors. After ten

kb in size. This insert was

G-protein-coupled receptors

peripheral blood leukocyte [a]

-27 to 625) of the rabbit F3B

... were isolated. The insert of

11

[REDACTED]

XX

```
b      134 -----IleLeuLeuAlaCysTlSerValAsnArgTrpLeuAlaIlovaluu 150
```

```

OY      894 TCTTCCACCTTATGAAACACACATTCTGCTACTGTTGACAGGGGACTATCTGATCAGAAAT 835
Db      151 AlATrTArG-----ThrLeuThrGlnLysArgTylLeuValLysPhe 164
OY      834 ATTTGT-----CTGTGACATCCAGTACTTAACCTGTTTAAAAAA 796
Db      165 IleCyLeuSerIleTrpGlyLeuSerLeuLeuLeuValLeuProValLeuLeuPheArg 184
OY      795 AAACACAGTCTTCAGATTAAGAAATTAGGGGCTCTAAATAGAGTGTATTAATGATAATGT 736
Db      185 ArgThrValTyrSerSerAsnValSerProLa-----CysTyrGluAsp----- 199
OY      735 CCTTTTTTTGGACAGAGTCTGTCTTTAAACAAACAAAGAAATGAGAGTCTTA 676
Db      200 -----MetGlyAsnAsnThrAlaAsnTrpArgMetLeu 210
OY      675 CTAACTGCTGTCCTCCAGAGGGATTCATAGAGAT----- 637
Db      211 LeuArgIleLeuProGlnSerPheGlyPheIleValProLeuLeuIleMetLeuPheCys 230
OY      636 -----CGTAAGAAAGAAACACACCTCCAGGGGAGAGCAGCATCTGGATT 592
Db      231 TyrGlyPheThrLeuArgThrLeuPheLysAlaHisMetGlyGlnLysHisArgLamEt 250
OY      591 CAGATAAAGTACCAGAAATATATTAGCTATCTTTTGCAGACATGCCCTTTCTGTGT 532
Db      251 ArgValIlePheAlaValValLeuIlePheLeuLeuCysTrp-LeuProTyrAsnLeuVal 270
OY      531 AACACTTCA-----GTCAATTCAGAAAGGCAATGTGTAGA 498
Db      270 IleuLeuAlaAspThrLeuMetArgThrGlnValIleGlnGlnIupThrCysGluArg 288

RESULT 11
AAG80121
ID      AAG80121 standard; protein; 360 AA.
AC      AAG80121;
DX      17-JAN-2002 (first entry)
DT      Human CXCR2 protein.
DE      Chemokine; tumour diagnosis; colorectal; prostatic; organ rejection;
KW      inflammation; autoimmunity disease; metastasis; bronchial asthma; lupus;
KW      chronic bowel inflammation; rheumatoid arthritis; cystostatic;
KW      antiinflammatory; antiaesthetic; immunosuppressive; dermatological;
KW      antirheumatic; antiarthritic.
OS      Homo sapiens.
XX      MO200172830-A2.
XX      PD      04-OCT-2001.
XX      PF      02-APR-2001; 2001WO-EP03708.
XX      PR      31-MAR-2000; 2000DE-1016013.
XX      PA      (IFPP-) IFF PHARM GMBH.
XX      PA      (FORS/) FORSMANN U.
XX      PI      Forssmann W, Adermann K, Heitland A, Spodsberg N;
XX      DR      WPI; 2001-626256/72.
XX      PT      Diagnostic agent containing two or more receptor-specific ligands,
XX      PT      useful for detecting tumors, inflammation etc., also therapeutic use of
XX      PS      ligand inhibitors -
XX      CC      Disclosure; Page 12; 26pp; German.

This invention describes a novel diagnostic agent (A) comprising at least

```

CC two different ligands (I) for receptors (II) that are implicated in
CC disease. (A) are used for the diagnosis of tumors (especially colorectal
CC or prostatic), organ rejection, inflammation and autoimmune diseases.
CC Also inhibitors of (I) are used therapeutically against tumors (and their
CC metastases), inflammation (particularly bronchial asthma or chronic bowel
CC inflammation), or autoimmune diseases (rheumatoid arthritis or lupus),
CC where the (cardio)vascular, lymphatic, respiratory, nervous, digestive,
CC endocrine, motor or urogenital systems or skin are affected, and bone
CC marrow diseases. The products of the invention are chemokine derivatives
CC which have cytostatic, antiinflammatory, antisthmatic,
CC immunosuppressive, dermatological, antihematic, antiathletic,
CC Chemokines act on specific tumor and inflammatory cells through a
CC costimulation of chemokine receptors (CR), which control migration and
CC proliferation of these cells. AA80045-AA80128 represent human chemokine
CC fragments used to illustrate the method of the invention.

XX

SQ Sequence 360 AA;

Alignment Scores:

Pred. No.:	Length:
Score: 0.344	Matches: 360
Percent Similarity: 85.50	Conservative: 58
Best Local Similarity: 37.84%	Mismatches: 40
Query Match: 22.39%	Indels: 87
	Gaps: 75
	Indels: 10

US-09-630-244B-2 (1-1331) x AAG80121 (1-360)

OY 1187 AGTCGATATTGGCATCAAGCCTCAATTAATCAATGCTTACTGATG-----ATA 1133

Dd 76 SerArgValGlyArgSerValThrSpValTYrLeuLeuAsnLeuAlaLeuValAspLeu 95

OY 1133 TTAAACGTAATATGTGTGTTGTTGGTGTTGATGTTGCCTAGTCAGTTCGTTCTCGG 1074

Dd 96 LeuPheAlaLeuThyrLeuProIleTrpAlaIleSerLysValAsnGlyTYrIlePheGly 115

OY 1073 AACCTGCGATGAAGGCTCTTTGGTTAAAGAACAATATCCAGAAATAGATATTAACATT 1014

Dd 116 ThrPheLeu-----CysLysValValSerLeuLeuLysGlyValAsnPh 130

OY 1013 GGAACCTAAGGGGATATAATGACACAGTAAAGGTGAAGGGTTTTTTTTCTTAAGGAG 954

Dd 131 -----TySerGly----- 133

OY 953 ACCAAGATTATTTATTTACTCAATCACTGCCA-AAAGCTTACTTTAAATGCTGTT 895

Dd 134 -----IleLeuLeuAlaLacysIleSerValIsparGlyrLeuAlaIleValHis 150

OY 894 TCTTCCCACTTTATGAAMCACCACTTTGCTACTGTGGACAGGGCACTTCGANTCAGAAAT 835

Dd 151 AlaThrArg-----ThreuthrgInLysArgTYrLeuValLysPhe 164

OY 834 ATTTGT-----CTGTTCAGATCAGACTTAATCTGTTAAAAAAA 796

Dd 165 IlecySLeuSerIleTrpGlyLeuSerLeuLeuLeuAlaLeuProValIleLeuPhearg 184

OY 795 AAACAGTGGCTTAGATAGCAATTAAGGGCTCTTAATAGATGTTTAAATGATTAATGT 736

Dd 185 ArgThryValTYrSerSerAsnValSerProAla-----CystyrGuLasp----- 199

OY 735 CCTTTTCTTTTGAGACAGAGTCTGTGCTTAAAAAACAAACAAAGAAATGAGAGCTTGA 676

Dd 200 -----MetGlyAsnAsnThrIlaAsnTrpArgmetLeu 210

OY 675 CTAACTGTCCTGTGCTCCAGAGGGATTCCAATGAGAGAT----- 637

Dd 211 LeuArgIleLeuPrgInsErPheGlyPheIleValrProLeuLeuIleMetLeuPheCys 230

OY 636 -----CGTAAGAAGGAAACAAACAACTCCAGGGAGCAAGCAGCATTTGGATT 592

Dd 231 TyrcIyPheThrLeuArgThrLeuPheLysAlaHisMetcIyGlnLysHisArgIlaMet 530

OY 591 CAGATAAAGTACCAGAAATATATAGTATCTCTTTCTGACAGCATGCGCTTTCTCTGTT 532

Db 251 ArgValIlePheAlaValIleuIlePheLeuLeuYsTrp-LeuProtyrAsnLeuVa 270
 QY 531 AACACTTCA-----GTCATTCAAAAGCGTATGCTGACA 498
 Db 270 IleuLeuAlaAspThrLeuMetArgThrGlnValIleGlnGluThrCysGluArg 288

RESULT 12
 AAU10557
 ID AAU10557 standard; Protein; 360 AA.
 XX
 AC AAU10557;
 XX
 DT 14-FEB-2002 (first entry)
 XX
 DE Human Interleukin 8 receptor beta (IL8RB) polypeptide.
 XX
 KM Human: Interleukin 8 receptor beta; IL8RB; antiinflammatory;
 KM haplotyping; haplotype pair; single nucleotide polymorphism; genotyping;
 KM gene therapy; drug screening; chronic obstructive pulmonary disease;
 KM inflammatory disease.
 XX
 OS Homo sapiens.
 XX
 PN MO200179221-A2.
 XX
 PD 25-OCT-2001.
 XX
 PF 12-APR-2001; 2001MO-US11942.
 XX
 PR 12-APR-2000; 2000US-196734P.
 XX
 PA (GENA-) GENA15555C50/07.
 XX
 PI Bentivegna SC, Chew A, Choi JY, Denton RR, Nandabalan K;
 XX
 DR N-PSDB; AAS16876.
 XX
 PT New polymorphic variants comprising Interleukin-8 receptor beta (IL8RB)
 PT isogene, useful in expressing IL8RB protein for use in screening for
 PT candidate drugs to treat diseases related to IL8RB activity, e.g.
 XX
 PS Inflammatory disorders
 XX
 PS Claim 29; Fig 3; 74pp; English.

The invention relates to single nucleotide polymorphisms in the human
 interleukin 8 receptor beta (IL8RB) gene. A method for haplotyping the
 IL8RB gene in an individual comprises identifying the nucleotide at one
 or more polymorphic sites and determining whether one of the copies of
 the gene is defined by one of the IL8RB haplotypes given in the
 specification or whether both copies are defined by a haplotype pair.
 This method is useful in genotyping, whereby all possible haplotype pairs
 can be assigned to specific genotypes. An association between a trait and
 a haplotype or haplotype pair of the IL8RB gene can be identified by
 comparing the frequency of the haplotype or haplotype pair in a
 population exhibiting the trait with the frequency of the haplotype or
 haplotype pair in a reference population, where a higher haplotype
 frequency in the trait population indicates the trait is associated with
 the haplotype or haplotype pair. IL8RB and its corresponding DNA are used
 for studying the expression and function of IL8RB, for use in screening
 for candidate drugs to treat diseases related to IL8RB activity, such as
 chronic obstructive pulmonary disease and other inflammatory disorders.
 The sequences are also useful for studying the effect of variation on the
 biological activity of IL8RB as well as on the binding affinity of
 candidate drugs targeting IL8RB. This sequence represents the IL8RB
 polypeptide.

Sequence 360 AA:

Alignment Scores: 0.344 Length: 360
 Pred. No.: 85.50 Matches: 58
 Score:

Percent Similarity: 37.84%
 Best Local Similarity: 22.39%
 Query Match: 3.70%
 DB: 23
 Gaps: 10

US-09-830-244b-2 (1-1331) x AAU10557 (1-360)

QY 1187 AGTCGATTTTGGCAATCAAGCCATCAATCAATGCTCGCTGATGATG-----ATA 1134
 Db 76 SerArgValGlyArgSerValThrAspValTyrLeuLeuAsnLeuAlaLeuAlaAspLeu 95
 QY 1133 TTAACACGATATTTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 1074
 Db 96 LeuPheAlaLeuThrLeuProIleThrAlaAlaSerLeuValAsnGlyThrPheGly 115
 QY 1073 AACTGCGATATAGGCTCTTTGGTAAAGCAATATCCAGAAATTAAGTAATTAATCTTT 1014
 Db 116 ThrPheLeu-----CysLysValValSerLeuLeuLysGluValAsnPhe 130
 QY 1013 GGAACTAAGGGGATATAATGACACAGTAAGGTGAAGGCTTTTCTTAAGGGAG 954
 Db 131 -----TyrSerGly----- 133
 QY 953 ACCAGAGTTTATTTATTAATCAATCAATGCTGCTCA-AAAAGTTACTTTAAATGCTGTTT 895
 Db 134 -----IleLeuLeuLeuAlaCysIleSerValAspArgTyrLeuAlaIleValHis 150
 QY 894 TCTTCCACATTTATGAAACACACATTTGCTACGCTGACAGGGGACTATCTGATCAGAANT 835
 Db 151 AlaThrArg-----ThrLeuThrGlnLysArgTyrLeuValLysPhe 164
 QY 834 ATTTGT-----CTGTGACATGACGATTAACGTGTTTAAAAAAA 796
 Db 165 IleCysLeuSerIleTyrGlyLeuSerLeuLeuAlaLeuProValLeuLeuPheArg 184
 QY 795 AAAACAGTCCCTGATGATGAAGAAATTTAGGGCTCTTAATCAATGCTTAAATGATATGT 736
 Db 185 ArgThrValTyrSerSerAsnValSerProLa-----CysTyrGluAsp----- 199
 QY 735 CTTTCTTTTGTGAGACAGAGTCTGCTTAAACAAACAAAGAAATGAGAGTCTA 676
 Db 200 -----MetLysAsnAsnThrAlaAsnThrPheMetLeu 210
 QY 675 CTAACTGCTCTGCTCTCCAGAGGGGATTCATAGAGAT----- 637
 Db 211 LeuArgIleLeuProGlnSerPheGlyPheIleValProLeuLeuIleMetLeuPheCys 230
 QY 636 -----CGTAAGAAAGAAACAACTCGACGGGACAGCAGCATCTTGATTT 592
 Db 231 TyrGlyPheThrLeuArgThrLeuPheLysAlaHisMetGlyGlnLysHisArgAlaMet 250
 QY 591 CAGTTAAAGTACCAAAAAATATAGCTATCTCTTTCAGACACTGCTCTTCTGTT 532
 Db 251 ArgValIlePheAlaValIleuIlePheLeuLeuYsTrp-LeuProtyrAsnLeuVa 270
 QY 531 AACACTTCA-----GTCATTCAAAAGCGTATGCTGACA 498
 Db 270 IleuLeuAlaAspThrLeuMetArgThrGlnValIleGlnGluThrCysGluArg 288

RESULT 13
 AAR70124
 ID AAR70124 standard; Protein; 1064 AA.
 XX
 AC AAR70124;
 XX
 DT 14-FEB-1996 (first entry)
 XX
 DE IL8-R type 2-GBP 130 fusion protein.
 XX
 KM Hybrid peptide; malaria parasite; Plasmodium falciparum; fusion protein;
 KM red blood cell; cytokine receptor; glycoporphin binding peptide 130;
 KM GBP 130; GBP; glycoporphin binding peptide homologue; glycoporphin A.

Hybrid peptides for binding cytokines, comprising a malaria parasite cell (Plasmodium falciparum) peptide (capable of binding to a red blood cell (RBC)) and a receptor peptide are claimed. AAR70103-25 are examples of these hybrid peptides. AAR70124 is a fusion of Interleukin 8 receptor type 2 and glycoporphin binding protein (GBP) 130. The use of cytokine receptors not normally found on RBCs means that the cytokine can bind harmlessly to the RBC without deleterious effect. The RBC protects the hybrid peptides from excretion from the kidney, and due to steric hindrance prevents the cytokines binding to a receptor in another cell. GBP 130 or GBPH (GBP homologue) are the pref. malaria parasite peptides used, others include EBA 175 (175 kDa erythrocyte binding antigen), PMW54 (pre major merozoite surface antigen) and the Duffy binding receptor molecule (eg. exhibited by Plasmodium vivax). These peptides bind to pref. glycoporphin A, B and C, sialo glycoproteins, found on the surface of RBCs. The hybrid peptides are thus used to lower the levels of free cytokines in the circulation to reduce pathological damage.

US-09-830-244B-2 (1-1331) x AAR70124 (1-1064)

YY	894	TCTTCCACGACTTTTGAAACCACATTGGTCATAGCTGTGCACGGGACATATCTGCATCAGAANT	835
DB	145	AATThArg-----ThLeThrGlnLysArgTrgYrIleValLysPhe	158
QY	834	ATTVTGT-----CGTTGACATMGAGTACTTAAGCTTTAAAAAAA	796
Db	159	IIECyLeuSerIleTrpPolyleuSerIleLeuAlaIeuProValLeuLeuPheArg	178
QY	795	AAAACGCTCCCTTCAGATAGAANAATTAGGGCCTTAATAAGATTGTTAATGATAATG	736
Db	179	ArgThrValTyrrSerSerAsnValSerProAla-----CystyrGluasp-----	193
QY	735	CCTTTTTTTTGAGACAGAGTCTCTGCTCAAAAAACAACAAGAAATGGAGAGTTCTA	676
Db	194	-----MetGlyAsnAsnThrAlaAsnTrpArgMetLeu	204
QY	675	CTAATGTCCTGTGCTCCACAGGGGANTTCAATAGAGAT-----	637
Db	205	IIEArgIleLeuProGlnSerPheGlyPheIleValProLeuLeuIleMetLeuPheCys	224
QY	636	-----CGTAAGACAGCAACAACCTGCAGGGGAAACAGCATCTTGGAATTT	592
Db	225	TyrglyPheThrIleuArgThrLeuPheLysAlaHisMetGlyGlnLysHisArgAlaMet	244
QY	591	CAGATAAGTAGTCCAAAAAATVTTAGCTATCTCTTCTGCAGACATGCTCTTCTGTGT	532
Db	245	ArgValIlePheAlaValIleuIlePheLeuLeuCysTrp-LeuProTyrrAsnLeuVala	264
QY	531	AACCTTTTCA-----GTCATTCAAAGGCTAATGGTAGA	498
Db	264	IIEuLeuAlaAspThrLeuMetArgThrGlnValIleIleGlnIthrCysGluArg	282
RESULT 14			
ID	AAB24943	standard; Peptide: 535 AA.	
XX	AAB24943;		
AC	27-NOV-2000	(first entry)	
DT	XX		
DE	Plant SDF encoded polypeptide sequence SEQ List 1 NO:474.		
KW	Plant; corn; Arabidopsis thaliana; sequence-determined DNA fragment;		
KM	SDF; genetic mapping; identification; promoter; structural gene; UTR;		
KX	untranslated region; expression control.		
OS	Plant.		
XX	MO2000040695-A2.		
PN	13-JUL-2000.		
PD	07-JAN-2000; 2000MO-USO0466.		
PF	08-JAN-1999; 99US-0115293.		
PR	(CERE) CERES INC.		
PA	Alexandrov N, Brover V, Chen X, Subramanian G, Troukhan ME;		
PI	Zheng L;		
PX	WPI: 2000-465970/40.		
DR	New corn plant and Arabidopsis thaliana sequence-determined DNA		
XX	fragments, useful for expressing gene products and for controlling		
PT	expression of a target gene -		
XX	Claim 14; Page 571-572; 673pp; English.		
XX	The present invention describes polynucleotides, such as complete cDNA		
CC	sequences and/or sequences of genomic DNA encompassing complete genes,		
CC	portions of genes, and/or intergenic regions, collectively referred		

The present invention describes polynucleotides, such as complete cDNA sequences and/or sequences of genomic DNA encompassing complete genes, portions of genes, and/or intergenic regions, collectively referred to

Pred. NO.:	0.468	Length:	53
Score:	85.00	Matches:	38
Percent Similarity:	44.06%	Conservative:	25
Best Local Similarity:	26.57%	Mismatches:	56
Query Match:	3.68%	Indels:	25
DB:	21	Gaps:	6

```

Db      |||:::|||||||CAGTACGATGATATAAAGCATATTGTGGTA-----111
134 LeuSerSerHisIleLeuAlaasnThrIleValPheLeuProTyrlLeuLeuIleAla 153
QY      1112 -----TTGGTGTCGATATGCCTTGGCGGTACTGCTGTTCGG 107#
Db      154 IleIleTyrservalSerLeutyrPheLeuValIGlyLeucysPheSerTrpIlnalau 173
QY      1073 AACTCGCGTAAAGGCTCTTGGTATTAAGAACAATATGCAGAAATTAAGATTAATCACTT 1014
Db      174 AlatyrrPhevalIleValIIetrPleIleValIleuMetAlaasnSerPheAlaleupe 193
QY      1013 GGAACATA-----GGCATATTAATGACACACAGTAAGGTGGA 976
Db      194 LeuSerSerLeuAlaIleProasnTyrlIleAlaGlyThrSerServalThIleLeuLeuAla 213
QY      975 GGATTTTTTTCTTAAGGAGACACAGTTTATATTACTCAAAATCAGTCCCAAAA 916
Db      214 AlaPhePheLeuPheSerGly-----tyrPheIleSerLysGluSerLeuProLys 230
QY      915 AGTTACTTTAAATGCTGTTTCTTCCACACTTTATGAACACACTTGGTACTGTTTACA 856
Db      231 TytTrPheuphemectyrrPhePheSermetyrylsynrAlalaLeuasralaleuIle 250
QY      855 GGGGACAT---CTGATCACAAATATATTGTCTGTGG-----ACATCAGTACT 811
Db      251 AsnGluytyrSerCyseuHisAsnLysCyseuLeuValIrPhegluGlnAlaSerValasn 270
QY      810 AACTGTTTA 802
Db      271 SerCyseu 273

```

XX	XXXX942;
XX	
DT	
DE	
XX	
XX	
DE	Plant SDF encoded polypeptide sequence SEQ List 1 NO:473.
XX	
KW	SDF; corn; Arabidopsis thaliana; sequence-determined DNA fragment;
KW	Pf1; genetic mapping; identification; promoter; structural gene; UTR
KW	untranslated region; expression control.
XX	
OS	Plant.
XX	
PN	WO200040695-A2.

Claim 14, Page 569-570: 673pp. English.

The present invention describes polynucleotides, such as complete cDNA sequences and/or sequences of genomic DNA encompassing complete genes, CC portions of genes, and/or intergenic regions, collectively referred to CC as sequence-determined DNA fragments (SDFs), from corn plants and CC *Arabidopsis thaliana*. The SDFs are promoters, structural genes, CC untranslated regions (UTRs), or 3' termination sequences. They can be CC used for expressing a gene product and controlling expression of a CC target gene, either as a promoter, a structural gene, an UTR or as a CC 3' termination sequence. They are also useful as tools for genetic CC mapping, and identification of a particular individual plant or for CC clustering a group of plants with a common trait. AAT78433 to AAT8630 CC and AAB24605 to AAB25099 represent the specifically claimed CC polynucleotide sequences and polypeptides encoded by them given in the CC present invention.

SQ Sequence 575 AA;

Pred. No.:	0.482	Length:	575
Score:	85.00	Matches:	38
Percent Similarity:	44.06%	Conservative:	25
Best Local Similarity:	26.57%	Mismatches:	56
Query Match:	3.68%	Indels:	6
DB:	21	Gaps:	25

US-09-830-244B-2 (1-1331) X AAB24942 (1-575)

[illegible]

Mon Jun 30 08:44:15 2003

us-09-830-244b-2.n2p.rag

Page 13

Db 291 AsnGluTyrSerCysLeuHisAsnLysCysLeuValITrPheGluGluAlaSerValAsn 310
Oy 810 AACTGTTTA 802
:::|||||
Db 311 SerCysLeu 313

Search completed: June 24, 2003, 19:19:50
Job time : 72 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 24, 2003, 19:17:37 ; Search time 18.5 Seconds
(without alignments)
4233.720 Million cell updates/sec

Title: US-09-830-244B-2
Perfect score: 2322
Sequence: 1 ctatgttttcagaatcaag.....gccttcacatgactctg 1331

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_plus_n2p -DEV=xlp
-Q=/cg22.1/USPNC_SPOOL/US09830244/runat_24062003.130002.10088/app_query.fasta.1.1479
-DB=Issued_Patents_AA -QFMT=fastn -SUFFIX=n2p.ra1 -MINMATCH=0.1 -LOOPCL=0
-LISTEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-MODE=LOCAL -OUTFMT=plco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -ALIGN=15
-USER=US09830244 -ECGN 1.1.28 -urnat_24062003.130002.10088 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEDQUERY -NEG_SCORES=0 -WAIT -DSPLLOCK=100 -LONGLOG
-DEV.TIMEOUT=10 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

- 1: Issued_Patents_AA:*
- 2: /cg22.6/ptodata/1/1aa/5A.COMB.pep:*
- 3: /cg22.6/ptodata/1/1aa/5B.COMB.pep:*
- 4: /cg22.6/ptodata/1/1aa/6A.COMB.pep:*
- 5: /cg22.6/ptodata/1/1aa/6B.COMB.pep:*
- 6: /cg22.6/ptodata/1/1aa/backfile1.pep:*

Pred: NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	93	4.0	384	4	US-09-071-035-276
2	93	4.0	430	4	US-09-071-035-276
3	85.5	3.7	360	1	US-08-202-056-7
4	85	3.7	2391	2	US-08-446-855A-2
5	85	3.7	2391	4	US-09-150-741-2
6	84.5	3.7	355	1	US-07-759-568-1
7	84.5	3.7	355	1	US-08-450-393A-8
8	84.5	3.7	355	2	US-08-390-000A-5
9	84.5	3.7	355	4	US-08-446-669-8
10	84.5	3.7	355	5	PCR-US95-00476-8
11	84	3.6	378	4	US-09-082-088-2
12	84	3.6	378	4	US-09-546-117-2

C	13	83.5	3.6	668	4	US-09-173-151A-35	Sequence 35, Appl
C	14	82.5	3.6	375	2	US-08-495-695B-28	Sequence 28, Appl
C	15	82.5	3.6	375	5	PCR-US94-14436-28	Sequence 18, Appl
C	16	79.5	3.4	2787	4	US-09-245-041-15	Sequence 15, Appl
C	17	77	3.3	615	4	US-09-134-001C-3284	Sequence 3284, Ap
C	18	76.5	3.3	350	1	US-08-202-056-1	Sequence 1, Appl
C	19	76.5	3.3	350	1	US-08-076-093A-2	Sequence 7, Appl
C	20	76.5	3.3	350	1	US-08-450-393A-7	Sequence 1, Appl
C	21	76.5	3.3	350	1	US-08-410-453A-1	Sequence 1, Appl
C	22	76.5	3.3	350	1	US-08-701-265-2	Sequence 2, Appl
C	23	76.5	3.3	350	1	US-08-410-454A-1	Sequence 1, Appl
C	24	76.5	3.3	350	2	US-08-284-586-2	Sequence 2, Appl
C	25	76.5	3.3	350	2	US-08-805-478-2	Sequence 2, Appl
C	26	76.5	3.3	350	2	US-08-802-627A-2	Sequence 2, Appl
C	27	76.5	3.3	350	2	US-08-801-238-2	Sequence 2, Appl
C	28	76.5	3.3	350	2	US-08-801-238-2	Sequence 2, Appl
C	29	76.5	3.3	350	2	US-09-104-296-2	Sequence 2, Appl
C	30	76.5	3.3	350	4	US-08-446-669-7	Sequence 7, Appl
C	31	76.5	3.3	350	5	PCR-US95-00476-7	Sequence 7, Appl
C	32	76.5	3.2	401	3	US-09-029-267-2	Sequence 2, Appl
C	33	74.5	3.2	350	2	US-08-495-695B-32	Sequence 32, Appl
C	34	73.5	3.2	451	4	US-08-495-484-12	Sequence 12, Appl
C	35	73.5	3.2	2165	5	PCR-US95-12507-2	Sequence 2, Appl
C	36	73	3.1	485	2	US-08-477-451-22	Sequence 22, Appl
C	37	73	3.1	490	4	US-09-126-420A-16	Sequence 16, Appl
C	38	72.5	3.1	366	4	US-09-134-001C-4412	Sequence 4412, Ap
C	39	72.5	3.1	156	4	US-09-134-001C-4412	Sequence 4, Appl
C	40	72.5	3.1	432	2	US-08-644-034A-2	Sequence 2, Appl
C	41	72	3.1	1226	2	US-08-540-804-12	Sequence 12, Appl
C	42	71.5	3.1	1226	2	US-08-218-265-12	Sequence 12, Appl
C	43	71.5	3.1				
C	44	71.5	3.1				
C	45	71.5	3.1				

ALIGNMENTS

RESULT 1
US-09-071-035-276
Sequence 276, Application US/09071035
Patent No. 6448043
GENERAL INFORMATION:
APPLICANT: Gil H. Choi
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
NUMBER OF SEQUENCES: 496
CORRESPONDENCE ADDRESS:
ADDRESS: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/071,035
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: A. Anders Brookes
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB369P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 276:
SEQUENCE CHARACTERISTICS:

LENGTH: 384 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-071-035-276

Alignment Scores:
Pred. No.: 0.00599
Score: 93.00
Percent Similarity: 36.26%
Best Local Similarity: 28.24%
Query Match: 4.02%
DB: 4
Matches: 74
Conservative: 21
Mismatch: 85
Indels: 83
Gaps: 12

US-09-830-244b-2 (1-1331) x US-09-071-035-276 (1-384)

672 ACTGCTCTGCTCCAGAGGGGATTCATGAGATGCTAAGAAAGAAACACTGCGAG 613
149 ThValLeuAlaAspLysGlnGlnGlnGlnPheAsn-----Gln 159
612 GGGAGAGAGCATCTGGATTTCAGATTAAG-----TAC 580
160 Gly-----GlnHisIleAsnTyrGlnLeuThrThrGlnIleProAlaAsnIleLeuGlyTyr 178
579 CAAAAAATTTAGCTATCTCTTTCGACAGATGCTCTTCTGTTACACTTTCAGT 520
179 Gln-GluPheAspLeuSerAspLysAlaAspThrThrLeuThrLeuLeuProGlnSerIle 198
519 CATTCAAAAGCGTATGAGTGAAG-----TCGACGCAAAATCCAGTG 479
198 egluValLysValAlaGlyLysThrValThrThrGlyTyrThrLeuThrThrGlnLysHis 218
478 TAGATTACTGTAAACAGTTTATTATA-----ACTATGCT 446
218 scLysPheThrLeuAspPheSerIleLysAspLeuGlnAsnPheAlaAsnGlnThrMetTh 238
445 GATATCA-----CCAATAAGCGCATTTACCAATTAAGCA-----412
238 rValSerTyrGlnMetAlaGlnGlnLysThrAlaGlnProAspThrAlaIleAsnAsnGln 258
411 -----TTAGTCGACATCAAAAGATGTAAGAAATGCTAATGGACATTTCCACAAAG 359
258 uGlyGlnLeuValThrAspLysHisThrLeuThrLysArgAlaThrValArgThrGlyGln 278
358 AAAG-----AAATCCAAATGCTCTCTATA-----336
278 LysSerPheValLysValAspSerGlnAsn-AlaLysIleThrLeuProGlnAlaValP 298
335 -----AAGGAGAAATATACCATGGCAACAGACAAATGAATTTCTGTC 293
298 heIleValLysAsnGlnAlaGlyIleTyrLeuAsnGlnThrAlaAsnGlyTyrArgTyrP 318
292 AAAAAGAAATTTAGCTATCTCTTTCGACAGATGCTCTTCTGTTACACTTTCAGT 520
318 LysGln-----LysAlaLeuAlaLysLysPheThrSerAsnGlnAlaGlyLys 334
232 CACATTAGCCACATATCTCAGATTAATTTGTTTACCTTAGGGCCCTCAATATTC 173
335 -----PheSerValLysGly**LysArgT 343
172 CCCCTATCTGCTAAACTTCACTGTTTACTCACAAGTGTGATTAATTTCCACCAAGTA 113
343 rProValLeuLeuGlnLysArgAsnLeuLysThrLysArgLeuSer-----SerGlnSer 361
112 ACCTTAACATTCCTTATGGCAGTGCAGAAATTTCCAGACATTTGTAACACTGTAATTTCA 53
361 snArgAsnSerPheTyrGlyGlyLysLysPheLeuLysAsnGlnLysArgThrThrAsnSer 381
52 CG 51
381 hr 381

RESULT 2

US-09-071-035-274
Sequence 274, Application US/09071035
Patent No. 6448043
GENERAL INFORMATION:
APPLICANT: Gil H. Choi
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
NUMBER OF SEQUENCES: 496
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/071,035
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: A. Anders Brooks
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB369P2
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8504
INFORMATION FOR SEQ ID NO: 274:
SEQUENCE CHARACTERISTICS:
LENGTH: 430 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-071-035-274

Alignment Scores:

Pred. No.: 0.00634
Score: 93.00
Percent Similarity: 36.26%
Best Local Similarity: 28.24%
Query Match: 4.02%
DB: 4
Matches: 430
Conservative: 74
Mismatch: 21
Indels: 85
Gaps: 12

US-09-830-244b-2 (1-1331) x US-09-071-035-274 (1-430)

672 ACTGCTCTGCTCCAGAGGGGATTCATGAGATGCTAAGAAAGAAACACTGCGAG 613
193 ThValLeuAlaAspLysGlnGlnGlnGlnPheAsn-----Gln 203
612 GGGAGAGAGCATCTGGATTTCAGATTAAG-----TAC 580
204 Gly-----GlnHisIleAsnTyrGlnLeuThrThrGlnIleProAlaAsnIleLeuGlyTyr 222
579 CAAAAAATTTAGCTATCTCTTTCGACAGATGCTCTTCTGTTACACTTTCAGT 520
223 Gln-GluPheAspLeuSerAspLysAlaAspThrThrLeuThrLeuLeuProGlnSerIle 242
519 CATTCAAAAGCGTATGAGTGAAG-----TCGACGCAAAATCCAGTG 479
242 egluValLysValAlaGlyLysThrValThrThrGlyTyrThrLeuThrThrGlnLysHis 262
478 TAGATTACTGTAAACAGTTTATTATA-----ACTATGCT 446
262 scLysPheThrLeuAspPheSerIleLysAspLeuGlnAsnPheAlaAsnGlnThrMetTh 282

```

OY 445 GATATCA-----CCAAATACGGCATTACCAATATATGCA-----417
Db 282 rValSerTyrGlnMetArgLeuGlnTyrSThrAlaGlnProAspThrAlaIleAsnAsnGln 302
OY 411 -----TTATGCGACATCAAAAGATCTGAAATGCTATATGGGACATTTCCACAAAAG 355
Db 302 uGlyGlnLeuValThrAspLysHisThrLeuThrLysAlaGlnAlaThrValAlaGlnGlyGln 322
OY 358 AAAG-----AAATCCATGGCTCCATA-----336
Db 322 YLysSerPheValLysValAspSerGlnAsn-AlaLysSrlLeuThrProGlnAlaValAla 342
OY 335 -----AAGGAGATATATAGCAATGAGCGACAGACAGATATAAATTTCTGTC 293
Db 342 heIleValLysAsnGlnAlaGlyGlnTyrLeuAsnGlnThrAlaAsnGlyTyrArgTrpG 362
OY 292 AAAAGAAAGATCTTCCAAAGCAATCAGAAAAAAGCTGTCCCAATATAGCGCTGGGAGC 233
Db 362 LmYsgGlu-----LysAlaLeuAlaLysLysPheThrSerAsnGlnAlaGlyGln- 378
OY 232 CACATTAGCCATATCTCAGATACGATTATTTTGTTTACGCTTAGGCGCCTCAATATTC 173
Db 379 -----PheSerValLysGly**LysArgT 387
OY 172 CCCCTATCTGCTAAACATTCGAATGTTTACTCACAGATCGATGAATATTCACCCAGTA 113
Db 387 rProValLeuLeuGlnArgAsnLeuGlySThrLysArgLeuSer-----SerGlnSerA 405
OY 112 ACCCTAAATTCCTCCATGGCGACGTGCAAAATTCACGACATTTGTAAACACTGTAAATTTCA 53
Db 405 snArgAsnSerPheTyrGlyGlnLysLysPheLeuGlyAsnGlnArgTrpThrThrAsnSerT 425
OY 52 CG 51
Db 425 hr 425

```

RESULT 3
 US-08-202-056-7
 Sequence 7, Application US/08202056
 Patent No. 5440021
 GENERAL INFORMATION:
 APPLICANT: Chuntcharapai, Anan
 APPLICANT: Hebert, Caroline
 APPLICANT: Kim, Kyung Jin
 APPLICANT: Lee, James
 TITLE OF INVENTION: Antibodies to Human IL-6
 NUMBER OF SEQUENCES: 8
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genentech, Inc.
 STREET: 460 Point San Bruno Blvd
 CITY: South San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94080
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 5.25 Inch, 360 Kb floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: patin (Genentech)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/202.056
 FILING DATE: 25-FEB-1994
 CLASSIFICATION: 436
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/677211
 FILING DATE: 29-MAR-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Love, Richard B.
 REGISTRATION NUMBER: 34,659
 REFERENCE/DOCKET NUMBER: 706P3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415/225-5530
 TELEFAX: 415/952-9881

```

;       TELEX: 910/371-7168
;       INFORMATION FOR SEQ ID NO: 7
;       SEQUENCE CHARACTERISTICS:
;           LENGTH: 360 amino acids
;           TYPE: amino acid
;           TOPOLOGY: linear
US-08-202-056-7

```

Alignment Scores:	
Pred. No.:	0.0534
Score:	85.50
Percent Similarity:	37.84%
Best Local Similarity:	22.39%
Query Match:	3.76%
DB:	1
Length:	36
Matches:	58
Conservative:	60
Mismatches:	67
Indels:	75
Gaps:	10

US-05-830-244B-2 (1-1331) X US-08-202-056-7 (1-360)

```

QY 1187 AGCCGATTTGGCCAAATCAACGCGCAATTAATCATGTCCTCCAGTACTGATG-----ATA 1133
Db |||||:::||||:::||||:::||||:::||||:::||||:::||||:::||||:::||||:::||||:::
QY 76 SerArgValGlyArgSerValThrAspValTyrLeuLeuAsnLeuAlaLeuAlaAspLeu 95
Db |||||:::||||:::||||:::||||:::||||:::||||:::||||:::||||:::||||:::
QY 1133 TTTAAACGATATATTGGTATATTGGCGTTGGATAGTGTGGCTTACTGGTGTTCGG 107
Db LeuPheAlaLeuThrLeuProIleThrPalaAlaSerLysValAsnGlyTyrIlePheGly 115
QY 1073 AACTGCGTGAATAAGCGCTTTGGTAAAGCAATATCCAGAAATATAGATATTAACTT 101
Db |||||:::||||:::||||:::||||:::||||:::||||:::||||:::||||:::||||:::
QY 116 ThrPheLeu-----CysLysValValSerLeuLeuLysGluValAsnPhe 130
Db |||||:::||||:::||||:::||||:::||||:::||||:::||||:::||||:::||||:::
QY 1013 GGAACTAAGGGATATAATGACACACAGTAAAGCTGGAAAGCGTTTTTTTCTTAAGGAG 954
Db |||||:::||||:::||||:::||||:::||||:::||||:::||||:::||||:::||||:::
QY 953 ACCAGAGTTTATATATTAATCAATACAGTCCTCCCA-AAAAGTTACTTTAAATGCTGTT 895
Db |||||:::||||:::||||:::||||:::||||:::||||:::||||:::||||:::||||:::
QY 134 -----IleLeuLeuLeuAlaCysIleSerValAspArgTyrLeuAlaIleValHis 150
QY 894 TCTTCCCACTTTATGAAACACACATTTGCTACTGTGGACAGGGACATCTCATGACGAAT 835
Db |||||:::||||:::||||:::||||:::||||:::||||:::||||:::||||:::||||:::
QY 151 AlaThrArg-----ThrLeuThrGlnLysArgTyrLeuValLysPhe 164
QY 834 ATTTGT-----CTGTGACATCAGTACTTAACGTGTTAAAAAA 796
Db |||||:::||||:::||||:::||||:::||||:::||||:::||||:::||||:::||||:::
QY 165 IleCysLeuSerIleTyrGlyLeuSerLeuLeuAlaLeuProValIleLeuPheArg 184
QY 795 AAACAGTGCCTTCAGATTAAGAAATTTAGGGGCTCTAATAGATGTATATGATAAATGT 736
Db |||||:::||||:::||||:::||||:::||||:::||||:::||||:::||||:::||||:::
QY 185 ArgThrValTyrSerSerAsnValSerProAla-----CysTyrGluAsp----- 199
QY 735 CCTTTTTTTGACAGACAGTCTCTGTTTAAAAAACAAGAAATAGAGAGTCTTA 676
Db |||||:::||||:::||||:::||||:::||||:::||||:::||||:::||||:::||||:::
QY 200 -----MetGlyAsnAsnThrAlaAsnTyrArgMetLeu 210
QY 675 CTAACGTCTGTCGTCCTCCAGAGGGAGTTCGAATGAGAT----- 637
Db |||||:::||||:::||||:::||||:::||||:::||||:::||||:::||||:::||||:::
QY 211 LeuArgIleLeuProGlnSerPheGlyPheIleValProLeuLeuIleMetLeuPheCys 230
Db |||||:::||||:::||||:::||||:::||||:::||||:::||||:::||||:::||||:::
QY 636 -----CGTAAGAAGAAACAAACACCTCCAGCGGACAGCATCTTGATTT 592
Db |||||:::||||:::||||:::||||:::||||:::||||:::||||:::||||:::||||:::
QY 231 TyrGlyPheThrLeuArgThrLeuPheLysAlaHisMetGlyLysHisAlaGlyMet 250
QY 591 CAGATAAGTACCAAAAATATATAGCATCTCTTTTGCACAGACATCGCTCTCTCTT 532
Db |||||:::||||:::||||:::||||:::||||:::||||:::||||:::||||:::||||:::
QY 251 ArgValIlePheAlaValIleLeuIlePheLeuLysCysTrp-LeuProTyrAsnLeuVal 270
QY 531 AACATTTTCA-----GTCATTCAAAAGGCGATTCGTAGA 498
Db |||||:::||||:::||||:::||||:::||||:::||||:::||||:::||||:::||||:::
QY 270 IleuLeuAlaAspThrLeuMetArgThrGlnValIleGlnGluThrCysGluArg 288

```

```

Patent No. 5849573
GENERAL INFORMATION:
APPLICANT: Stewart, Thomas S
APPLICANT: Flores, Maria V
APPLICANT: O'Sullivan, William J
TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl
NUMBER OF INVENTION: phosphate synthetase II
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon & Vanderhye PC
STREET: 1100 No. 5849573th Glebe Road, 8th Floor
CITY: Arlington
STATE: Virginia
COUNTRY: USA
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446,855A
FILING DATE: 06-Jul-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mitchard, Leonard C
REGISTRATION NUMBER: 29,009
REFERENCE/DOCKET NUMBER: 47-80
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-816-4000
TELEFAX: 703-816-4100
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2391 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-446-855A-2

Alignment Scores:
Pred. No.: 0.159
Score: 85.00
Percent Similarity: 38.85%
Best Local Similarity: 26.11%
Query Match: 3.68%
DB: 2
Gaps: 18

US-09-830-244b-2 (1-1331) x US-08-446-855A-2 (1-2391)
QY 930 ATCAGTCTCCCAAAAAGTACTTAAATGCTGTTTCTCCCACTTATGAACCAT 871
DB 572 IISerLeuGlyCysAspThrTyrIysMetIySTyrgIyAsnIyGlyValasngIInPro 591
QY 870 TTGGCACTTTGACAGGGGACATATGATCAGAAATATTTGTCGTGACATGACTT 811
DB 592 ValIle-----GlnLeuValaspaSnIleCysTyrIleThrSerGlaasn 606
QY 810 AAC-----TGTAAAAAANAACAGCGCTTCAGATGAATTAAGGGGCTCTATA 757
DB 607 HIsGlyTyrcysLeuLeuIySylSerIleLeuIySAtyrgIyGluLeuAlaIleSerTyr 626
QY 756 ACATGTTTATGATAAATGTCCTTTTGTGAGCAGAGCTCTGCTTAAAAACAAA 697
DB 627 IleaSnAlaAsnAspLysSer-Ile-----GlnGlyIleSer----- 638
QY 696 ACAAGAATAGGAGAGTCTTACTACTGCTGCTGCTCCAGAGGATTCAT--GA 641
DB 639 -HIsYsAsnGlyAspThrTyrSerValGlnPheHIsProGluIySAsnGlyProG 658
QY 640 GCAT-----CGTAAGAAGCAAAACAA 620
DB 658 uAspThrSerPheLeuPheIySAsnPheLeuAspIlePheAsnIySylSyls----- 676

```

```

QY 619 CCTGAGGGGAGAGCATCTTGATTTACGATAAGTAC-----CAAAAAATATTAGC 566
DB 677 ----GlnTyrArgGlyTyrLeuGlyTyrAsnIleIleTyrIleIySylSylValIle 695
QY 565 TATCTCTTTCGACACATGCTCTTCTCTGTTAACTGATTCAGTCAAAAGCTA 506
DB 695 uLeuGlySerGlyIyLeuGlyIleGlyGlnAlaGlyIubPheAspTyrSer---GlyTh 714
QY 505 TTGGTAGAGGTCGACCAATCCAACTAGATTTACTGTAACTATTAATTAATGCT 446
DB 714 GlnAlaIleIySylSerLeuIySylGlyCys-----GlyIleTyrValIleLeuVa 730
QY 445 GTATTCACCAATTAACGGCATTTACCAATATGTCATTTAGTCATCAAAAGATCTGAAA 386
DB 730 IAsnProAsnIleAlaThrVal-Gln-----ThrSerIyGlyLeuAla 745
QY 385 TGCTAATGGGACATTTCCACAAAGAAAGAAATCCAAATGGCTCTRTAAAGGAGAA 326
DB 745 sPlySValTyrPheLeuPheValasncysGluPheValGlyIySylIleIySylSyl 765
QY 325 ATAGCAAT-----GGCAAGCAGAAATGAATTTCTGT----- 294
DB 765 ySProAspPheIleLeuGlyThrPheGlyIyGlnIrrAlaLeuAsncysAlaLeuMet 785
QY 293 -----CAAAAGAAAGTACTTCTCAAGCAATCAAGAAAGTCTGCCAATTAAGGCTG 239
DB 785 euAspGlnIySylValIleuIySyls-----AsnAsncysGlnCysLeuGlyThr- 801
QY 238 GGGAGCCACATTTAGCCACTATTCACATGACTATTTGTTAGGCTTAGGGGCTCA 179
DB 802 -----SerLeuGlnSerIleA 807
QY 178 ATATTCCTCCCATCTGCTAAGTCACTGTTTACTCAGAGTCTGATTAATTCACAC 119
DB 807 rGlieThrGlu-----AsnArgThrLeuPhe-----AlaGlyIySylLeuIySg 821
QY 118 CACGTAACCTTAACAT--CCCTATGGCAGTCCAAA 84
DB 821 IuIleAsnGluArgIleAlaProTyrGlySerAlaIySyls 833

RESULT 5:
US-09-150-741-2
: Sequence 2, Application US/09150741
: Patent No. 6183996
: GENERAL INFORMATION:
: APPLICANT: Stewart et al.
: TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
: Patent No. 6183996
: FILE REFERENCE:
: CURRENT APPLICATION NUMBER: US/09/150,741
: CURRENT FILING DATE: 1998-09-10
: EARLIER APPLICATION NUMBER: PL6380
: EARLIER FILING DATE: 1992-12-16
: EARLIER APPLICATION NUMBER: AU93/00617
: EARLIER FILING DATE: 1993-12-02
: EARLIER APPLICATION NUMBER: 08/446,855
: EARLIER FILING DATE: 1995-07-06
: NUMBER OF SEQ ID NOS: 15
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 2
: LENGTH: 2391
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: protein
US-09-150-741-2

Alignment Scores:
Pred. No.: 0.159
Score: 85.00
Percent Similarity: 38.85%
Length: 2391
Matches: 82
Conservative: 40

```

Best Local Similarity: 26.118
Query Match: 3.688
DB: 4
Mismatches: 110
Indels: 83
Gaps: 18

US-09-830-244B-2 (1-1331) x US-09-150-741-2 (1-2391)

```

QY 930 ATCAGTCTCCCAAAAAGTACTTTAAATGCTGTTTCTCCCACTTATGAAACCAT 871
DB 572 IIESErLeuGlyCyAspThrTyrLysMetLysTyrGlyAsnArgGlyValAsnGlnPro 591
QY 870 TTGCTACGTGTACAGGGGACATCTGATCAGAAATATTTCTGTTGACATCAGTACTT 811
DB 592 ValIle-----GlnLeuValAspAsnIleCysTyrIleThrSerGlnAsn 606
QY 810 AAC-----TGTTTAAAAAAAACAGTGCCTCAGATAGAAATTAAGGGCTCTATA 757
DB 607 HIsGlyTyrCysLeuLysLysLysSerIleLeuLysArgLysGlnLeuAlaIleSerTyr 626
QY 756 AGATGTTTAAATGATTAATGCTCTTTTGTGAGACAGAGTCTCTGTTTAAACAA 697
DB 627 ILeAsnAlaAsnAspLysSer-Ile-----GlnGlyLysSer----- 638
QY 696 ACAAGAAATGAGAGGTTCTACTACTGCTCTGCTGCTCCAGAGGGGATTCAT---GA 641
DB 639 -HisLysAsnGlnLysArgPheTyrSerValGlnPheHisProGlnGlyAsnGlnProGln 658
QY 640 GGAT-----CGTAAAGAGAAACAA 620
DB 658 uAspThrSerPheLeuPheLysAsnPheLeuLeuAspIlePheAsnLysLysLys----- 676
QY 619 CCTGAGGGAGAGACATCTTGATTTTCAGATTAACATAC-----CAAAATATTAGC 566
DB 677 -----GlnTyrArgGlnLysLeuGlyTyrAsnIleIleTyrIleLysLysValLeuLeu 695
QY 565 TATCTCTTTCTGAGACATGCTCTTCTCTGTTAACACTTTCAGATTCATCAAAAGCTTA 506
DB 695 uLeuGlySerGlyGlyLeuLysLysIleGlyGlnAlaGlyLysLysLysSer---GlyIle 714
QY 505 TTGGTAGAGGTGACAGCAATTCAGATGATTAAGTACTGTAACAGTATTATTAAGTACTGCT 446
DB 714 rGlnAlaIleLysSerLeuLysGlnLys-----GlyIleTyrValIleLeuVal 730
QY 445 GTATTACCAATTAAGCGCATTAACCATTAATGCGATTAAGCGCATCAAAAGATCTGAAA 386
DB 730 LAsnProAsnIleAlaThrVal-Gln-----ThrSerLysGlyLeuAla 745
QY 385 TGCTAATGGACATTTCCCAAAAGAAAGAAATCCATGCTGCTCTATTAAGGGAGAT 326
DB 745 sPlyValTyrPheLeuProValAsnGlyGlnPheValGlnLysIleIleLysLysGlnL 765
QY 325 ATAGCAAT-----GGCAGCAGATGAAATTTCTGT----- 294
DB 765 ySerProAspPheIleLeuLysThrPheGlyGlnAlaThrAlaLeuAsnGlyAlaLeuMetL 785
QY 293 -----CAAAAGAAAGTACTTCTCAAAAGCAATGAGAAACAGTCTCCCAATTAAGCCTG 239
DB 785 euAspGlnLysLysValLeuLysLys-----AsnAsnGlyLysLysLysLysLysThr- 801
QY 238 GGGAGCCACATTAAGCCACTATCTCAGATAGTATGATTTGTTTGGCTTGGGGCCCA 179
DB 802 -----SerLeuGlnSerIleA 807
QY 178 ATATTCCTCCCTCATCTGCTAAACTCAACTGTTTACTCAGACAGTCTGATTAATTCAGC 119
DB 807 rGlnThrGln-----AsnArgThrLeuPhe-----AlaGlnLysLeuLysG 821
QY 118 CAAGTAACTTAACAT---CCCTATGAGCAGTGAAGAA 84
DB 821 LulleAsnGlnLysGlnLeuLeuProTyrGlySerAlaLys 833

```

```

: Patent NO. 5374506
: GENERAL INFORMATION:
: APPLICANT: Murphy, Philip M.
: TITLE OF INVENTION: Cloning of cDNA Encoding a Functional
: TITLE OF INVENTION: Human Interleukin-8 Receptor
: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Cushman, Darcy & Cushman
: STREET: 1615 L Street, N.W.
: CITY: Washington
: STATE: D.C.
: COUNTRY: U.S.A.
: ZIP: 20036-5601
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/759,568
: FILING DATE: 19910913
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Scott, Watson T.
: REGISTRATION NUMBER: 26581
: REFERENCE/DOCKET NUMBER: WTS/5683/91535/MBH
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-861-3000
: TELEFAX: 202-822-0944
: TELEX: 6714627 cush
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 355 amino acids
: TYPE: AMINO ACID
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: US-07-759-568-1
:
: Alignment Scores:
: Pred. No.: 0.0713 Length: 355
: Score: 84.50 Matches: 58
: Percent Similarity: 37.84% Conservative: 40
: Best Local Similarity: 22.39% Mismatches: 87
: Query Match: 3.65% Indels: 75
: DB: 1 Gaps: 10
:
: US-09-830-244B-2 (1-1331) x US-07-759-568-1 (1-355)
:
: QY 1187 AGTCGATTTCGCCAATCAACCTCAATATCATGCTGCGACTTACTGATG---ATA 1134
: DB 71 SerArgValGlyArgSerValThrAspValTyrLeuLeuAsnLeuAlaLeuAspLeu 90
:
: QY 1133 TTTAAAGTATATTGTTGATGATGTTGCTGATGTTGCTGATGTTGCTGATGTTGCTG 1074
: DB 91 LeuPheAlaLeuThrLeuProIleThrPalaAlaSerLysValAsnGlyTyrPheGly 110
:
: QY 1073 AACTGCTGATAAGGCTCTTGGTATAAAGCAATATCCAGAAATPAGATATTAATTT 1014
: DB 111 ThrPheLeu-----CysLysValValSerLeuLeuLysLysLysValAsnPro 125
:
: QY 1013 GGAACCTAAGGGGATTAATATGACACAGTAAGGTGGAAGGCTTTTCTTAAGCGAG 954
: DB 126 -----TyrSerGly----- 128
:
: QY 953 ACCAGATTATTAATTACTCAATCAGTCCCA-AAAAGTTACTTTAAATGCTGTTT 895
: DB 129 -----IleLeuLeuLeuAlaCysIleSerValAspArgTyrLeuAlaIleValHis 145
:
: QY 894 TCTCCCACTTATATGAACACACATTTGCTACTGTTGACAGGGGACTATCTGATCAGAAAT 835
: DB 146 AlaThrArg-----ThrLeuThrGlnLysArgTyrLeuValLysPhe 159

```

RESULT 6
US-07-759-568-1
: Sequence 1, Application US/07759568

QY 834 ATTGT-----CTGTGACATGACTTAACTGTTAAAAA 796
 |||||
 Db 160 Ilecysleuserilertpelyleuserleuleleuvalleuprovalleuleuphearg 179
 QY 795 AAAACAGTCCCTTCAGATAGAAATTAGGGCTTAATAGATTATGATTAATGT 736
 |||||
 Db 180 ArgthrvalltyrserSerSernValSerProAla-----CystylcIuasp----- 194
 QY 735 CTTTTTTTTAGACAGAGTCTGTCTTTAAAAAACAAACAAAGAAATGAGACTTCTA 676
 |||||
 Db 195 -----MetGlyAsnAsnThrAlaAsnThrArgMetLeu 205
 QY 675 CTAACTGCTGCTGCTCCAGAGGGGATTCATGAGAT----- 637
 |||||
 Db 206 LeuArgilleleuprogInserphegilypheilevalproleuleullemetleuphecys 225
 QY 636 -----GTAAGAGGAAACAAACACTGCAGGGGAGGAGCAGACAGCTTGGATT 592
 |||||
 Db 226 Tyrglypethrleuargthrleupheylsalahismetglylnlyshisargalamet 245
 QY 591 CAGATAAGTACCAAAAAATATTATCTATCTCTTTCGACAGACATGCTCTTCTCTGT 532
 |||||
 Db 246 Argvalillephealavalvalleulleuleucystrip-leuprotlyrAsnleuva 265
 QY 531 AACACTTCA-----GTCATTCAAAAGGCTATTGGTAGA 498
 |||||
 Db 265 lleuleuAlaAspThrleuMetargThrglnvalilleglnluthrcysgluarg 283

RESULT 7

US-08-450-393A-8
 : Sequence 8, Application US/08450393A
 : Patent No. 5707815

: GENERAL INFORMATION:
 : APPLICANT: Chairo, Israel
 : APPLICANT: Coughlin, Shaun
 : TITLE OF INVENTION: MAMMALIAN MONOCYTE CHEMOATTRACTANT
 : TITLE OF INVENTION: PROTEIN RECEPTORS
 : NUMBER OF SEQUENCES: 14
 : CORRESPONDENCE ADDRESS:
 : ADDRESS: Cooley Godward Castro Huddleson & Tatum
 : STREET: 5 Palo Alto Square
 : CITY: Palo Alto
 : STATE: California
 : COUNTRY: USA

: ZIP: 94306-2155
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.25

: CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/450,393A
 : FILING DATE: May 25, 1995
 : CLASSIFICATION: 424
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Coert, Luan

: REGISTRATION NUMBER: 31,822
 : REFERENCE/DOCKET NUMBER: UCAL-237/020US
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 415-843-5165
 : TELEFAX: 415-8857-0663
 : TELEX: 380816COOLEYPA

: INFORMATION FOR SEQ ID NO: 8:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 355 amino acids
 : TYPE: amino acid

: TOPOLOGY: linear
 : MOLECULE TYPE: protein
 : HYPOTHEICAL: NO
 : US-08-450-393A-8

Alignment Scores:

Pred. No.: 0.0713

Length: 355

Score: 84.50
 Percent Similarity: 37.84%
 Best Local Similarity: 22.39%
 Query Match: 3.65%
 DB: 1
 Gaps: 10

US-09-830-244B-2 (1-131) x US-08-450-393A-8 (1-355)

QY 1187 AGTCGATTATTCATCAAGCCTCAATATCATGCTGCTGATGATG-----ATA 1134
 |||||
 Db 71 SerArgValIlglyrSerValThrAspValIlyrleuleuAsnleuAlaAspIleu 90
 QY 1133 TTAACGATATATTTGTTGCTATTTGTTGCTGATGTTGCTGCTGATGCTTCTGCG 1074
 |||||
 Db 91 LeuphealaleuthrleuproiletrpAlaIalaserlyValAsnlylTrrIleheclly 110
 QY 1073 AACTGCTGATTAAGGCTCTTGGTATTAACCATATTCACAAATATTAATTAATCTT 1014
 |||||
 Db 111 Thrheleu-----CylslyValIalSerleuleuLysgluValAsnph 125
 QY 1013 GGAACATAAGGGATTAATGACACAGTAAGCTGAAGGGCTTTTCTTAAGGAG 954
 |||||
 Db 126 -----Tyrsergly----- 128
 QY 953 ACCAGATTATTAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 895
 |||||
 Db 129 -----lleuleuleuAlaCysIleSerValAspArglyrleuAlaIleValHis 145
 QY 894 TCTTCCACATTTATGAACACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 835
 |||||
 Db 146 AlathrArg-----ThrleuThrArglyrleuValIlysphe 159
 QY 834 ATTGT-----CTGTGACATGACTTAACTGTTAAAAA 796
 |||||
 Db 160 Ilecysleuserilertpelyleuserleuleuvalleuprovalleuleuphearg 179
 QY 795 AAAACAGTCCCTTCAGATAGAAATTAGGGCTTAATAGATTATGATTAATGT 736
 |||||
 Db 180 ArgthrvalltyrserSerSernValSerProAla-----CystylcIuasp----- 194
 QY 735 CTTTTTTTTAGACAGAGTCTGTCTTTAAAAAACAAACAAAGAAATGAGACTTCTA 676
 |||||
 Db 195 -----MetGlyAsnAsnThrAlaAsnThrArgMetLeu 205
 QY 675 CTAACTGCTGCTGCTCCAGAGGGGATTCATGAGAT----- 637
 |||||
 Db 206 LeuArgilleleuprogInserphegilypheilevalproleuleullemetleuphecys 225
 QY 636 -----GTAAGAGGAAACAAACACTGCAGGGGAGGAGCAGACAGCTTGGATT 592
 |||||
 Db 226 Tyrglypethrleuargthrleupheylsalahismetglylnlyshisargalamet 245
 QY 591 CAGATAAGTACCAAAAAATATTATCTATCTCTTTCGACAGACATGCTCTTCTCTGT 532
 |||||
 Db 246 Argvalillephealavalvalleulleuleucystrip-leuprotlyrAsnleuva 265
 QY 531 AACACTTCA-----GTCATTCAAAAGGCTATTGGTAGA 498
 |||||
 Db 265 lleuleuAlaAspThrleuMetargThrglnvalilleglnluthrcysgluarg 283

RESULT 8

US-08-390-000A-5
 : Sequence 5, Application US/08390000A
 : Patent No. 5985583

: GENERAL INFORMATION:

: APPLICANT: Sealton, Stuart C.

: TITLE OF INVENTION: Cloning and Expression of

: NUMBER OF SEQUENCES: 8

: CORRESPONDENCE ADDRESS:

: ADDRESS: Pennie & Edmonds LLP

: STREET: 1155 Avenue of the Americas

: CITY: New York


```

Db      119 PheValAlaLeuClAlaSerThrCysSerLeuLeuAlaIleAlaIleGluArgHisLeu 138
      |||
      119 ThMetLeuIleuSerIlePheThrAlaLeuValThrIleValIleLeuTyAlaA 217
      :|:
      139 ThrMetLeuIleuSerIlePheThrAlaLeuValThrIleValIleLeuTyAlaA 217
      :|:
      872 -----TGtGGTTTCATPAAAGTGGGAAGAAACAGCATTTTAAAGTACTTTTGGGACA 925
      |||
      159 GlyMetCysTrpLeuIleAlaPheThrIleuGluAlaLeu-ProIleLeuGlyTTrpAsnC 178
      :|:
      926 CT-----CATTTGAGTAAATATAAACTGTGGTCTCCCTTAAGAAAAAAAACCTTCC 979
      :|:
      178 yslEuHisInleuProAspCysSerThrIleLeuProleuTySerIlyS--TyrI 197
      :|:
      980 ACCTTACAGTGTGC--ATTATATCCCTTAGTTCGCAAGTAATATATCTTATTTCTGG 1033
      :|:
      197 leAlaPheCysIleSerIlePheThrAlaIleuValThrIleValIleLeuTyAlaA 217
      :|:
      1037 ATATTGCTTTTATACCAAGACCTTATACGCCAGTCCAGAACACACATPACGCACA 1099
      |||
      217 rglElyrPheLeuValIlySerSerSerArgIyValAlaAsn-----HisA 233
      :|:
      1097 ACCATTACCAACCAATACCAACATATACGTTTAAATATCATC 1140
      |||
      233 snAsnSerGluArgSerMetAlaLeuArgThrValValIle 247

RESULT 12
US-09-546-117-2
/ Sequence 2, Application US/09546117
/ Patent No. 6344542
/ GENERAL INFORMATION:
/ APPLICANT: TSUI, PING
/ TITLE OF INVENTION: HUMAN EDG3b GENE
/ FILE REFERENCE: GP-70453-D1
/ CURRENT APPLICATION NUMBER: US/09/546,117
/ CURRENT FILING DATE: 2000-04-10
/ PRIOR APPLICATION NUMBER: 09/082,088
/ PRIOR FILING DATE: 1998-05-20
/ NUMBER OF SEQ ID NOS: 2
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 2
/ LENGTH: 378
/ TYPE: PRT
/ ORGANISM: HOMO SAPIENS
US-09-546-117-2

Alignment Scores:
Pred. No.: 0.0853 Length: 378
Score: 84.00 Matches: 41
Percent Similarity: 42.35% Conservative: 42
Best Local Similarity: 20.92% Mismatches: 69
Query Match: 3.62% Indels: 44
Gaps: 9

US-09-830-244B-2 (1-1331) x US-09-546-117-2 (1-378)
QY 653 CTCGTGGAGCAGACAGAGTGTAGTAACCTCTGCATTTCTTGGT----- 697
Db 64 lIeTrpYsAsnInlySPheHISAsnArgMetIlyrPheHleIleGlysnLeuAlaLeu 83
QY 698 -----TTGTTTAAAGACAGACACTCTGTCTCAAAAAAGAGCAT 739
Db 84 CysAspLeuLeuAlaGlyIleAlaTyrlYsValAsnIleuMetSerGlyIlySlyThr 103
QY 740 TTATCATTTAAACATCTTTATAGAGCCCTAAT-----TTCTATCTGAAGCACTGTT 793
Db 104 PheSerLeu-----SerProThrValTrpPheLeuArgGluGlySerMet 118
QY 794 TTTTAAAAACAGTTAGTACT-----GATGTCAACAGCAAAATA 835
Db 119 PheValAlaLeuClAlaSerThrCysSerLeuLeuAlaIleAlaIleGluArgHisLeu 138

```

QY 836 TTTCGATCAGATAGTCCCTGTCACAGTACCAAA----- 871
Db 139 ThMeIleuysMecharProtyrAspAlaenLysArgHisArgValPheLeuLeuIle 158
QY 872 -----TGTCGTTTCATTAAGTGGAGAGAAACAGCATTTTAACTATTGTTGGGGA 925
Db 159 GlyMeCysArgPheLeuIleAlaPheHsrLeuGlyAlaLeu--ProIleuGlyTyrPnc 178
QY 926 CT-----GATTGAGTATTAATAAACTGTCGTCCTTAAGAAAAAAACCTTCC 979
Db 178 yLeuHisnLeuProAspCysSerThrIleLeuProLeuTyrSerLysLys--TyrI 197
QY 980 ACCTTTACTGTCTC--ATTATATCCCTAGTCCAAATTAATTTATTTCTGCG 1036
Db 197 LeAlaPheCysIleSerIlePheHsrAlaIleLeuValThrIleValIleLeuTyrAla 217
QY 1037 ATATTGCTTTTATACCAAGACCTTATACAGCTTCAGACACACACATATACGACA 1096
Db 217 rGlyLeuTyrPheLeuValLysSerSerSerArgLysValAlaAsn-----HisA 233
QY 1097 ACCATACCAACCAATACCAACAAATACCTTTTATATATC 1140
Db 233 snAsnSerGluArgSerMetAlaLeuLeuArgThrValIle 247

RESULT 13
US-09-173-151A-35
Sequence 35, Application US/09173151A
Patent No. 6326472
GENERAL INFORMATION:
APPLICANT: Timans, Jacqueline C.
APPLICANT: Debels, Johannes Eduard Maria
APPLICANT: Antonius
APPLICANT: Sana, Theodore R.
APPLICANT: Bazan, J. Fernando
APPLICANT: Kastelein, Robert A.
TITLE OF INVENTION: Human Receptor Proteins; Related Reagents and Methods
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: DNA Research Institute
STREET: 901 California Avenue
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/173,151A
FILING DATE: 14-OCT-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/065,776
FILING DATE: 17-NOV-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/078,008
FILING DATE: 12-MAR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/081,883
FILING DATE: 15-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/095,987
FILING DATE: 10-AUG-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/078,416
FILING DATE: 18-MAR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/062,066
FILING DATE: 15-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Ching, Edwin P.

REGISTRATION NUMBER: 34,090
REFERENCE/DOCKET NUMBER: DX0767X
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650)852-9196
TELEFAX: (650)496-1200
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 668 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-173-151A-35
Alignment Scores:
Pred. No.: 0.131
Score: 83.50
Percent Similarity: 33.838
Best Local Similarity: 22.898
Query Match: 3.61%
DB: 4
Gaps: 11
Length: 668
Matches: 46
Conservative: 22
Mismatch: 50
Indels: 83

US-09-830-244B-2 (1-1331) x US-09-173-151A-35 (1-668)
QY 1155 TGTCTCGAGTACTGATGATTAATAACGTATATTGTTGGTATGTTGATGTT 1096
Db 3 CysThrAspTyrPheSerIleAspIleLysLysTyr-----GlnValLeuValGlyGluPro 20
QY 1095 GTGCGTATAGTGGTGTCTG-----GAACTGGCTGAT 1063
Db 21 ValArgIleLysCysAlaLeuPheHsrLeuTyrGlyTyrIleArgThrAsnTyrSerLeuAlaGln 40
QY 1062 AAGGCT-----CTTGGTATTAAGCAATATCCAGAAATAGATTAATTAAT 1015
Db 41 SerAlaGlyLeuSerLeuMetIleTyrLysSer----- 52
QY 1014 TGGAACTAAGGGCATATTAATGACACAGTAAGGTGAAGGCTTTTCTTAAAGCA 955
Db 53 ---GlyProGlyAspPheGluGluProIleAlaPheAspGlySerArgMetSerLysGlu 71
QY 954 GACCAAGTTTTATTATTACTCAATCAGTCTCCAAAGTAAGTTACTTAATAGCTGTT 895
Db 72 GluAspSerIle----- 75
QY 894 TCTTCCACTTATGAACACATTTGCTACTGTTGACAGGAGCTATGATCAGAAAT 835
Db 76 -----TriPheArgProThrLeuLeuGlnAspSerLysLeuTyr----- 88
QY 834 ATTGTCTGTGACATCAGTACTTAACTGTTTAAA-----AAAAACAGTG--- 787
Db 89 AlaCysValIleArgAsnSerThrTyrCysMetLysValSerIleSerLeuThrValGly 108
QY 786 CTTTCAGATAGAATAGGAGGCTTAATAGATTAATAGATTAATAGATTAATAGATTAAT 727
Db 109 GluAsnAspThrGlyLeu-----CysTyrAsnSerLysMetLysTyrPhe 123
QY 726 TGAGACAGAGCTCTGCTTAAATAAAACAA----- 697
Db 124 ---GluLysAlaGluLeuSerLysSerLysGluIleSerCysArgAspIleGluAspPhe 142
QY 696 -----ACAAAGAAATG 685
Db 143 LeuLeuProThrArgGluProGluIleLeuTyrTyrLysGluCysArgThrLysThrTyr 162
QY 684 AGA 682
Db 163 Arg 163
RESULT 14
US-08-495-695B-28
Sequence 28, Application US/08495695B
Patent No. 5976814
GENERAL INFORMATION:
APPLICANT: Bard, Jonathan A.

```

OY 380 TTACCATTTTCAGATCTTTTGATGTGCACATAAG-----CCATTAATTGGTAATGCCGTT 43
Db 83 LeuAlaIheSerAspIheLeuMetCys-LeuIleCysGlnProIleuThrValThrTyrTh 10
OY 434 ATTGGTGATATACGCACTACTGTAATAAATCAGTAAGCTGTACAGTAATCTCAGTGGATTTGGTG 49
Db 102 rIleMetAspTyrTrpIlephedgIValIleuLeuGlySerMetLeuThrPheIleGlnCy 12
OY 494 CACCTCTC-----ACCAATAGCCTTTTGAATGACTGAAAGTGTTAACAGAGAAAGAGCATGT 55
Db 122 smetSerValThrValSerIleLeuSer----- 13
OY 551 CTCGACAAAGAGATAGCAATATATTTTGGTACTTATCTGAAATCCCAAGATGCTGCTTC 61
Db 132 -----LeuValLeuValAlaIleuGlnIuArgHisGlnLeuIleIleAs 145
OY 611 CCTGCAGGTTGTTTTCCTTCTTTCAGATCCATCATGATCCCTCTGCGAGACAGACAGA 67
Db 145 nPThrArgIlyTrpIlyProSer-----IleSerGlnAlaTyrIleuGlyI 160
OY 671 GTTAGAGAACTCTCCATTTCTTTGTTTGTATTTTAAAGACAGAGACTGTCTCAAAA 73
Db 160 eValValIleTrpPheIleSer-----CysPheLeuSerLeuProPheLeuAlaIleAsn 178
OY 731 A-----AAGACATTTATCATTTATTAACATCTTATTAGAGCCCTAATTTCTTATCTGAA 784

```

```

Db          |||.....|||...::|||..|||..|||
   178 rIleLeuAsnspheNhenstYrAas--HisserLySValVaIgLnPheLau---Glu 196
Qy      GGCACSTGTATTTTAAACAAGTTAACTACTCATTCATCCCTTAGTCCAAGAATTATC 844
           |||    ||| ..:::||| ...||| ....
Db     AspLYStAlVAlCySPHeVAISerTPSerSerASphSIshArGLueULeTY----- 214
Qy       845 AGATAgTGCCCCCGTCCAACAGTAACAATGGCTTGCATAAGGGGAAGAAAACAGcAtt 904
        214 -----
Qy       905 TTAAGTAActTTTTGGGAGAcTAcTtGaGTAtATAAAcTcygTCccCTTAga 964
        214 -----
Db         215 -----ThrrHrrPhelEuLeu----PheGIInTyrcYSval 225
Qy       1025 TCttAtTTcgatTaTtcctTTATTCSCAAAGAccTTATcAGcAGTTCcGA----- 1076
                ::||||:::: | | | | |
Db       226 ProLeUalARhelLeuVALCYstyMeTarGlIEtyGlnArLeuslhnrgslndArg 245
               :| | | | | | | | | | |
Qy       1079 -----ACAAACCacTfATagGCCAACaccAfRccaAACCaAAATACcaAAATATATAC 1126
                ||| | | | | | | | |
Db       246 ARgalARPheNIshTrHIshTRCySSerSetarGVaIGlyglMetELypProILeanSlY 265
                ||| | | | | | | | |
Qy       1127 GTTTTAATATgcATgcATACSTcCACAGAcATGATTTATGAGCGTTGATTGGCAAATACcAc 1166
                ::|||::::: ||| | | | |
Db       266 MetLeumetalmeVelIThrLa----- 273
Qy       1187 TTC-----TAcATCcAAATTCCTCAATCTTTCATACCAATACAcACTTACCAC- 1234
                ||| ::::: :: :| | | | | | | | | |
Db       274 PheaIAvaLLeuTripleuProleuhIsValPheasnThrLeuglaApRTPrTyrlngLU 293
Qy       1235 -----TTTTgtTAGATcATCTPAAGAcCAANT- 1261
        1262 ---CGAA TGTA AACCTPATTA ATTACTG ATACTC 1294
                 ||||| | | | | | | |
Db       314 SerTHrcysValaspPropheLIETryClYPheLEu 325
                               ||||| | | | | | |
RESULT 15
PCF-US94-14436-28
Sequence 28, Application PC/TUS9414436
GENERAL INFORMATION:
APPLICANT: Bard, Jonathan A.
APPLICANT: Walker, Mary
APPLICANT: Branchek, Theresa
APPLICANT: Weinshtark, Richard L.
TITLE OF INVENTION: DNA ENCODING A HUMAN NEUROPEPTIDE
TITLE OF INVENTION: Y/PEPTIDE YY/PANCREATIC POLYPEPTIDE
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 1185 Avenue Of The Americas
City: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/M5-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCF/US94/14436
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
```

REGISTRATION NUMBER: 28, 678
 REFERENCE/DOCKET NUMBER: 44743-A-PCIT\JPM\MAT
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 278-0400
 TELEFAX: (212) 391-0525
 INFORMATION FOR SEQ ID NO: 28:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 375 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US94-14436-28

Alignment Scores:
 Pred. No.: 0 132
 Score: 82.50 Length: 375
 Percent Similarity: 34.13% Matches: 78
 Best Local Similarity: 23.35% Conservative: 36
 Query Match: 3.55% Mismatches: 102
 DB: 5 Gaps: 118
 14

US-09-830-244b-2 (1-1331) x PCT-US94-14436-28 (1-375)

```

QY 380 TRAGCATTTTCAGATCTTTGATGCGACTAATG-----CCATTAATGGTAATGCCGTT 433
DB 83 LeuAlaPheSerAspPheLeuMetCys-LeuIleCysGlnProLeuThrValThrTyTh 102
QY 434 ATTGGTAATACAGCATAGTTAATAAATCACTGTAACGTAATCTACACTGGATTGCTG 493
DB 102 rIleMetAspTyrTrIlePheGlyGlnValLeuCysLysMetLeuThrPheIleGlnCy 122
QY 494 CACCTCT--ACCAATAGCCTTTGATGACTGNAAGTGTAAACAGAGAAGAGCATGT 550
DB 122 smetSerValThrValSerIleLeuSer----- 131
QY 551 CTGCAGAAAGATACATAATATTTTGGTACTTATGCTGAATCCAGATGCTGCTTC 610
DB 132 -----LeuValLeuValAlaLeuGlnArgHisGlnLeuIleLeas 145
QY 611 CCCTGAGGTGTTTCTCTTACGATCCATGAAATCCCTCGGAGCAGCAGACA 670
DB 145 nProThrGlyTrpLysProSer-----IleSerGlnAlaTyLeuGlyI 160
QY 671 GTTAGTAGAAGCTCTCATTTCTTTGTTGTTTAAAGACAGAGACTGTCTCAAAA 730
DB 160 eValValIleTrpPheIleSer-----CysPheLeuSerLeuProPheLeuAlaAsnSe 178
QY 731 A-----AAGACATTTATCATATATACATCTTATAGAGCCCTAATTTCTTATCTGAA 784
DB 178 rIleLeuAsnAspLeuPheHisTyrAsn--HisSerLysValValGlnPheLeu---Gln 196
QY 785 GGCACGTGTTTTTTTAAACAGTTAAGTATGATGTCACAGACAGAAATATTTCTGATC 844
DB 197 AspLysValValCysPheValSerTrpSerSerAspHisHisArgLeuIleTyr----- 214
QY 845 AGATAGTCCCTGTCACACAGTAGCAAAATGTGTTTCATATAAGTGGAGAAACAGCATT 904
DB 214 ----- 214
QY 905 TTAAGTAATCTTTTGGAGACTGATTGAGTAATAATAAACTGTGTCCTCCTTAAGA 964
DB 214 ----- 214
QY 965 AAAAAAACCTTCACCTTACCTGATGATTTATATCCCTTATGTTCCAAAGTTAATTA 1024
DB 215 -----ThrThrPheLeuLeu-----PheGlnTyrCysVal 225
QY 1025 TCTTATTTCTGGATATCTTTATACCAAGAGCCTTATCAGCCACTTCCACA----- 1078
DB 226 ProLeuAlaPheIleLeuValCysTyMetArgIleTyGlnArgLeuGlnArgGlnArg 245
QY 1079 -----ACAACCCTATAGCGACCAACATACCAACACCAACATACCAACATATATAC 1126
DB 1079 ----- 1126

```

```

DB 246 ArgAlaPheHisThrHisThrCysSerSerArgValGlyGlnMetLysProIleAsnGly 265
QY 1127 GTTTAATATATCTATCAGTAAGTACGACAGACATGATTAATGAGCTTGATGGCAATACGAC 1186
DB 266 MetLeuMetAlaMetValThrAla----- 273
QY 1187 TTC-----TACATCCATATCTCATCTTTGATACCATATACACTACTACTACCAC--- 1234
DB 274 PheAlaValLeuTrpLeuProLeuHisValPheAsnThrLeuGlnAspTrpTyGlnGln 293
QY 1235 -----TTTGTGATCATCTTAAGCAATG--- 1261
DB 294 AlaIleProAlaCysHisGlyAsnLeuIlePheLeuMetCysHisLeuPheAlaMetVal 313
QY 1262 ---CGAATGTAAACCCCTATATATTTACTGATCTC 1294
DB 314 SerThrCysValAsnProPheIleTyrGlyPheLeu 325

```

Search completed: June 24, 2003, 19:26:12
 Job time : 33.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: June 24, 2003, 19:19:58 ; Search time 49.5 Seconds

(without alignments) 5819.122 Million cell updates/sec

Title: US-09-830-244B-2

Perfect score: 2322
Sequence: 1 catgtttttagaatacaag.....gcctttccaatgactcttg 1331

Scoring table:

BLOSUM62	Xgapop 10.0	Xgapext 0.5
Xgapop 10.0	Xgapext 0.5	
Xgapop 6.0	Xgapext 7.0	
Delop 6.0	Delext 7.0	

Searched: 417779 seqs, 108206813 residues

Total number of hits satisfying chosen parameters: 835558

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame_plus.n2p model -DEV=xip
-O=/cgn2_1/USPTO.spool/US09830244/Runat_24062003_130003_10114/app_query.fasta_1.1479
-DB=Published.Applications_AA -OPM=fastest -SUFFIX=n2p.rapb -MINMATCH=0.1
-LOOCL=0 -LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=PCT -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0
-NCPU=6 -ICPU=3 -NO_MMAP -LARGESUBSTR -NEG_SCORES=0 -WAIT -DSBLOCL=100
-LONGLOC -DEV.TIMEOUT=130 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

Published Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	150	6.5	78	9	US-09-924-340-6
2	150	6.5	78	9	US-09-992-600A-6
3	150	6.5	78	9	US-10-000-489-6
4	150	6.5	78	9	US-10-000-986-6

C	5	87.5	3.8	355	9	US-10-237-563-35	Sequence 35, Appl
C	6	85	3.7	1610	9 <td>US-10-155-533-9</td> <td>Sequence 9, Appl1</td>	US-10-155-533-9	Sequence 9, Appl1
C	7	84.5	3.7	355	9 <td>US-10-237-563-27</td> <td>Sequence 27, Appl</td>	US-10-237-563-27	Sequence 27, Appl
C	8	84.5	3.7	355	9 <td>US-10-237-563-28</td> <td>Sequence 28, Appl</td>	US-10-237-563-28	Sequence 28, Appl
C	9	84.5	3.6	433	10 <td>US-09-841-132-363</td> <td>Sequence 363, App</td>	US-09-841-132-363	Sequence 363, App
C	10	84	3.6	378	10 <td>US-09-971-228-7</td> <td>Sequence 7, Appl1</td>	US-09-971-228-7	Sequence 7, Appl1
C	11	84	3.6	378	10 <td>US-09-842-316-7</td> <td>Sequence 18, Appl</td>	US-09-842-316-7	Sequence 18, Appl
C	12	84	3.6	378	10 <td>US-09-731-030A-18</td> <td>Sequence 21, Appl</td>	US-09-731-030A-18	Sequence 21, Appl
C	13	84	3.6	378	12 <td>US-10-037-610-21</td> <td>Sequence 35, Appl</td>	US-10-037-610-21	Sequence 35, Appl
C	14	84	3.6	606	9 <td>US-09-876-790-4</td> <td>Sequence 29, Appl</td>	US-09-876-790-4	Sequence 29, Appl
C	15	83.5	3.6	668	9 <td>US-10-011-548-35</td> <td>Sequence 10112, A</td>	US-10-011-548-35	Sequence 10112, A
C	16	82.5	3.6	355	9 <td>US-10-237-563-29</td> <td>Sequence 15, Appl</td>	US-10-237-563-29	Sequence 15, Appl
C	17	79.5	3.4	1329	10 <td>US-09-815-242-10112</td> <td>Sequence 32, Appl</td>	US-09-815-242-10112	Sequence 32, Appl
C	18	79.5	3.4	2787	10 <td>US-09-893-228-15</td> <td>Sequence 362, App</td>	US-09-893-228-15	Sequence 362, App
C	19	79	3.4	426	10 <td>US-09-841-132-573</td> <td>Sequence 4, Appl1</td>	US-09-841-132-573	Sequence 4, Appl1
C	20	78.5	3.4	355	9 <td>US-10-237-563-32</td> <td>Sequence 30, Appl</td>	US-10-237-563-32	Sequence 30, Appl
C	21	78	3.4	412	10 <td>US-09-841-132-362</td> <td>Sequence 31, Appl</td>	US-09-841-132-362	Sequence 31, Appl
C	22	78	3.4	1332	9 <td>US-10-041-856-4</td> <td>Sequence 37, Appl1</td>	US-10-041-856-4	Sequence 37, Appl1
C	23	77.5	3.4	355	9 <td>US-10-237-563-30</td> <td>Sequence 83, Appl</td>	US-10-237-563-30	Sequence 83, Appl
C	24	77.5	3.4	355	9 <td>US-10-237-563-31</td> <td>Sequence 33, Appl</td>	US-10-237-563-31	Sequence 33, Appl
C	25	77.5	3.4	678	9 <td>US-09-895-913A-4</td> <td>Sequence 10, Appl</td>	US-09-895-913A-4	Sequence 10, Appl
C	26	77	3.3	355	9 <td>US-10-237-563-37</td> <td>Sequence 10, Appl</td>	US-10-237-563-37	Sequence 10, Appl
C	27	76.5	3.3	350	9 <td>US-10-237-563-37</td> <td>Sequence 4, Appl1</td>	US-10-237-563-37	Sequence 4, Appl1
C	28	76.5	3.3	350	9 <td>US-09-104-063-2</td> <td>Sequence 2, Appl1</td>	US-09-104-063-2	Sequence 2, Appl1
C	29	76.5	3.3	350	10 <td>US-09-782-980-83</td> <td>Sequence 7, Appl1</td>	US-09-782-980-83	Sequence 7, Appl1
C	30	76	3.3	350	10 <td>US-09-884-430-7</td> <td>Sequence 33, Appl</td>	US-09-884-430-7	Sequence 33, Appl
C	31	76	3.3	423	10 <td>US-10-237-563-33</td> <td>Sequence 8, Appl1</td>	US-10-237-563-33	Sequence 8, Appl1
C	32	76	3.3	1362	10 <td>US-09-808-483-8</td> <td>Sequence 14009, A</td>	US-09-808-483-8	Sequence 14009, A
C	33	75	3.2	1534	9 <td>US-09-736-968A-10</td> <td>Sequence 10, Appl</td>	US-09-736-968A-10	Sequence 10, Appl
C	34	75	3.2	1534	10 <td>US-09-736-960-10</td> <td>Sequence 10, Appl</td>	US-09-736-960-10	Sequence 10, Appl
C	35	75	3.2	1534	10 <td>US-09-815-242-14009</td> <td>Sequence 6, Appl1</td>	US-09-815-242-14009	Sequence 6, Appl1
C	36	75	3.2	2066	9 <td>US-09-978-244A-6</td> <td>Sequence 8, Appl1</td>	US-09-978-244A-6	Sequence 8, Appl1
C	37	75	3.2	2107	9 <td>US-09-815-379-8</td> <td>Sequence 5, Appl1</td>	US-09-815-379-8	Sequence 5, Appl1
C	38	74.5	3.2	621	9 <td>US-10-179-046-2</td> <td>Sequence 5331, Ap</td>	US-10-179-046-2	Sequence 5331, Ap
C	39	74.5	3.2	651	10 <td>US-09-331-631A-5</td> <td>Sequence 12682, A</td>	US-09-331-631A-5	Sequence 12682, A
C	40	74.5	3.2	651	10 <td>US-09-815-242-5331</td> <td>Sequence 1, Appl1</td>	US-09-815-242-5331	Sequence 1, Appl1
C	41	74.5	3.2	651	10 <td>US-09-815-242-12682</td> <td>Sequence 3, Appl1</td>	US-09-815-242-12682	Sequence 3, Appl1
C	42	74.5	3.2	666	9 <td>US-09-331-631A-1</td> <td>Sequence 34, Appl</td>	US-09-331-631A-1	Sequence 34, Appl
C	43	74.5	3.2	666	9 <td>US-10-237-563A-3</td> <td>Sequence 2, Appl1</td>	US-10-237-563A-3	Sequence 2, Appl1
C	44	74	3.2	355	9 <td>US-10-237-563-34</td> <td></td>	US-10-237-563-34	
C	45	74	3.2	1215	9 <td>US-10-232-539-2</td> <td></td>	US-10-232-539-2	

ALIGNMENTS

RESULT 1
US-09-924-340-6
; Sequence 6, Application US/09924340
; Publication No. US20030027248A1
; GENERAL INFORMATION:
; APPLICANT: Benjamin, Stephane
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91.US2.REG
; CURRENT APPLICATION NUMBER: US/09/924,340
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: JPatent
; SEQ ID NO 6
; LENGTH: 78
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: 1..19

US-09-924-340-6

Alignment Scores:

Pred. No.:	1.85e-08	Length:	78
Score:	150.00	Matches:	33
Percent Similarity:	80.95%	Conservative:	1
Best Local Similarity:	78.57%	Mismatches:	8
Query Match:	6.46%	Indels:	0
DB:	9	Gaps:	0

US-09-830-244B-2 (1-1331) x US-09-924-340-6 (1-78)

QY 1204 CATCTTTCATACATACACTACTACACTTTTGTNAGATCATCTAAGAGCAATGCG 1263

DB 37 HSHSSTYRHSILetHrleuLeuProleupheglugluserSerlysserAsnala 56

QY 1264 AATGTAAACCCATATATTAATTAATGATACCTTTGGTTCAGATACCTTTCCTTCCAAAT 1323

DB 57 ASnglulysHslyrAsnleuLeuTyThrleucysPheargileuAlaPheSerile 76

QY 1324 GTCACT 1329

DB 77 ValThr 78

RESULT 2

US-09-992-600A-6

Sequence 6, Application US/09992600A

Publication No. US20030027161A1

GENERAL INFORMATION:

APPLICANT: Benjamin, Stephanie

APPLICANT: Tanaka, Hiroaki

TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF

FILE REFERENCE: 91. US4. DIV

CURRENT APPLICATION NUMBER: US/09/992,600A

PRIOR FILING DATE: 2001-11-13

PRIOR APPLICATION NUMBER: US 09/924,340

PRIOR FILING DATE: 2001-08-06

PRIOR APPLICATION NUMBER: PCT/IB01/01715

PRIOR FILING DATE: 2001-08-06

PRIOR APPLICATION NUMBER: US 60/305,456

PRIOR FILING DATE: 2001-07-13

PRIOR APPLICATION NUMBER: US 60/302,277

PRIOR FILING DATE: 2001-06-29

PRIOR APPLICATION NUMBER: US 60/298,698

PRIOR FILING DATE: 2001-06-15

PRIOR APPLICATION NUMBER: US 60/293,574

PRIOR FILING DATE: 2001-05-25

NUMBER OF SEQ ID NOS: 114

SOFTWARE: JPatent

SEQ ID NO 6

LENGTH: 78

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: SIGNAL

LOCATION: 1..19

US-09-992-600A-6

US-09-830-244B-2 (1-1331) x US-09-992-600A-6 (1-78)

QY 1204 CATCTTTCATACATACACTACTACACTTTTGTNAGATCATCTAAGAGCAATGCG 1263

DB 37 HSHSSTYRHSILetHrleuLeuProleupheglugluserSerlysserAsnala 56

QY 1264 AATGTAAACCCATATATTAATTAATGATACCTTTGGTTCAGATACCTTTCCTTCCAAAT 1323

DB 57 ASnglulysHslyrAsnleuLeuTyThrleucysPheargileuAlaPheSerile 76

QY 1324 GTCACT 1329

DB 77 ValThr 78

RESULT 3

US-10-000-489-6

Sequence 6, Application US/10000489

Publication No. US20030092011A1

GENERAL INFORMATION:

APPLICANT: Benjamin, Stephanie

APPLICANT: Tanaka, Hiroaki

TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF

FILE REFERENCE: 91. US6. DIV

CURRENT APPLICATION NUMBER: US/10/000,489

PRIOR FILING DATE: 2001-11-14

PRIOR APPLICATION NUMBER: US 09/924,340

PRIOR FILING DATE: 2001-08-06

PRIOR APPLICATION NUMBER: PCT/IB01/01715

PRIOR FILING DATE: 2001-08-06

PRIOR APPLICATION NUMBER: US 60/305,456

PRIOR FILING DATE: 2001-07-13

PRIOR APPLICATION NUMBER: US 60/302,277

PRIOR FILING DATE: 2001-06-29

PRIOR APPLICATION NUMBER: US 60/298,698

PRIOR FILING DATE: 2001-06-15

PRIOR APPLICATION NUMBER: US 60/293,574

PRIOR FILING DATE: 2001-05-25

NUMBER OF SEQ ID NOS: 112

SOFTWARE: JPatent

SEQ ID NO 6

LENGTH: 78

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: SIGNAL

LOCATION: 1..19

US-10-000-489-6

US-09-830-244B-2 (1-1331) x US-10-000-489-6 (1-78)

QY 1204 CATCTTTCATACATACACTACTACACTTTTGTNAGATCATCTAAGAGCAATGCG 1263

DB 37 HSHSSTYRHSILetHrleuLeuProleupheglugluserSerlysserAsnala 56

QY 1264 AATGTAAACCCATATATTAATTAATGATACCTTTGGTTCAGATACCTTTCCTTCCAAAT 1323

DB 57 ASnglulysHslyrAsnleuLeuTyThrleucysPheargileuAlaPheSerile 76

QY 1324 GTCACT 1329

DB 77 ValThr 78

RESULT 4

US-10-000-986-6

Sequence 6, Application US/10000986

Publication No. US20030096247A1

GENERAL INFORMATION:

APPLICANT: Benjamin, Stephanie

APPLICANT: Tanaka, Hiroaki

TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF

FILE REFERENCE: 91. US9. DIV

CURRENT APPLICATION NUMBER: US/10/000,986

Assignment Scores:	
Read, No.:	0.526
Score:	87.50
Percent Similarity:	37.88
Percent Local Similarity:	22.788
Query Match:	3.788
Length:	355
Matches:	59
Conservative:	39
Mismatches:	87
Indels:	75

NUMBER OF SEQ ID NOS: 12
SOFTWARE: PatentIn version 3.1
SEQ ID NO 9
LENGTH: 1610
TYPE: PRT
ORGANISM: Plasmodium falciparum

QY 636-GTAAGAAGAAACACCTGAGGAGGACGATCTGGATT 592
 DB 226 TygIyphenrthleuArgrthleuPheLysAlaHisMetClyGlnHisArgAlaMet 245
 QY 591 CAGATAAGTACCAAAAAATATTAGCTATCTCTTTCTGCAGACATGCGCTCTTCTCTGT 532
 DB 246 ArgValIlePheAlaValIleuIlePheLeuLeuCysTrp-LeuProTyrAsnLeuVa 265
 QY 531 AACACTTCA-GTCAATCAAAAGCTATTGGTAGA 498
 DB 265 lleuLeuAlaAspThrleuMetArgThrGlnValIleGlnIuThrCysGluArg 283
 RESULT 8
 US-10-237-563-28
 ; Sequence 28, Application US/10237563
 ; Publication No. US20030082660A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Robert Horlick, Jinguao Zhao, Robert Swanson, Maria Webb, Barbara Strohl,
 ; TITLE OF INVENTION: Orthologues of Human Receptors and Methods of Use
 ; FILE REFERENCE: 1073.058AWO
 ; CURRENT APPLICATION NUMBER: US/10/237,563
 ; PRIOR FILING DATE: 2002-12-09
 ; PRIOR APPLICATION NUMBER: 09/576,160
 ; NUMBER OF SEQ ID NOS: 63
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 28
 ; LENGTH: 355
 ; TYPE: PRT
 ; ORGANISM: Pan troglodytes
 US-10-237-563-28
 Alignment Scores:
 Pred. No.: 1.17 Length: 355
 Score: 84.50 Matches: 38
 Percent Similarity: 37.84% Conservative: 40
 Best Local Similarity: 22.39% Mismatches: 87
 Query Match: 3.65% Indels: 75
 Gaps: 10
 US-09-830-244b-2 (1-1331) x US-10-237-563-28 (1-355)
 QY 1187 AGTCGATATTTGCCAATCAACCCCTCAATTAATCATGTCCTCAGTACTGATG-----ATA 1134
 DB 71 SerArgValAlaGlyArgSerValThrAspValTyrleuLeuAsnLeuAlaLeuAlaAspLeu 90
 QY 1133 TTAACAGCTATATTTGGTATTTGGTATTTGGTATTTGGTATTTGGTATTTGGTATTTGGT 1074
 DB 91 LeuPheAlaLeuThrleuProIleThrAlaAlaSerLysValAsnGlyTrpIlePheCly 110
 QY 1073 AACTGCTATATAGGCTCTTTGGTATTAAGCAATATCCAGAAATTAAGTATTAATCTT 1014
 DB 111 ThrPheLeu-----CysLysValAlaSerLeuLeuLysGlnValAsnPhe 125
 QY 1013 GGAACTAAGGGGATATTAATGACACAGTAAGGCTGAAGGCTTTTCTTAAGGAG 954
 DB 126 -----TyrSerGly----- 128
 QY 953 ACCAGAGTTTATTTATCTCAATCACTCTCCCA-AAAAGTTACTTTAAATGCTGTT 895
 DB 129 -----IleLeuLeuLeuAlaCysIleSerValAspArgTyrleuAlaIleValHis 145
 QY 894 TCTTCCCACTTTATGAACCAACATTTGCTACTGTGACAGGGGACTATTCGATCAGAAT 835
 DB 146 AlaThrArg-----ThrleuThrGlnLysArgTyrleuValLysPhe 139
 QY 834 ATTTGT-----CTGTGACATCAGTACTTAACTTTAAAAAAA 796
 DB 160 IleCysLeuSerIleTyrPglYleuSerLeuLeuAlaLeuProValIleuLeuPheArg 179
 QY 795 AAAACAGTGCCTTCAATGAATTAAGGAGCTCTTAATAGATGTTATATGATTAATAGT 736

DB 180 ArgThrValTyrSerSerAsnValSerProAla-----CysTyrGlnAsp----- 194
 QY 735 CTTTTTTTTGAGACAGAGCTCTGCTTCTTAAACAAACAAAGAAATGACAGTTCTA 676
 DB 195 -----MetGlyAsnAsnThrAlaAsnTrpArgMetLeu 205
 QY 675 CTAACTGCTCTGCTCCCGACAGGAGGATTCATGAGAT----- 637
 DB 206 LeuArgIleLeuProGlnSerPheGlyPheIleValIleProLeuLeuIleMetLeuPheCys 225
 QY 636-GTAAGAAGAAACACCTGAGGAGGACGATCTGGATT 592
 DB 226 TygIyphenrthleuArgrthleuPheLysAlaHisMetClyGlnHisArgAlaMet 245
 QY 591 CAGATAAGTACCAAAAAATATTAGCTATCTCTTTCTGCAGACATGCGCTCTTCTCTGT 532
 DB 246 ArgValIlePheAlaValIleuIlePheLeuLeuCysTrp-LeuProTyrAsnLeuVa 265
 QY 531 AACACTTCA-GTCAATCAAAAGCTATTGGTAGA 498
 DB 265 lleuLeuAlaAspThrleuMetArgThrGlnValIleGlnIuThrCysGluArg 283
 RESULT 9
 US-09-841-132-363
 ; Sequence 363, Application US/09841132
 ; Patent No. US20020061848A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bhatia, Ajay
 ; APPLICANT: Skelky, Tahir A.W.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
 ; FILE REFERENCE: 210121.469C8
 ; CURRENT APPLICATION NUMBER: US/09/841,132
 ; NUMBER OF SEQ ID NOS: 599
 ; SOFTWARE: FastSeq for Windows Version 3.0/4.0
 ; SEQ ID NO 363
 ; LENGTH: 433
 ; TYPE: PRT
 ; ORGANISM: Chlamydia
 US-09-841-132-363
 Alignment Scores:
 Pred. No.: 1.25 Length: 433
 Score: 84.50 Matches: 63
 Percent Similarity: 35.44% Conservative: 38
 Best Local Similarity: 22.11% Mismatches: 90
 Query Match: 3.64% Indels: 95
 Gaps: 14
 US-09-830-244b-2 (1-1331) x US-09-841-132-363 (1-433)
 QY 662 CACAGGACAGTATGTAACCTCCATTTCTTTGTTTCTTTTAAAGACA----- 712
 DB 6 HisHisMetTyrValArgSerIlePhePheSerIleIleAlaIlePheLeuThrValGlyCys 25
 QY 713 -----GAGACTGTG 721
 DB 26 SerPheSerProGlnSerIleLeuIleIleAlaIleHisAspAspProArgSerLeu 45
 QY 722 TCTCAAAAAAAGGA-----CATTTATATTAACATCTATTAAGAGCCCT 769
 DB 46 SerProGlnLysGlyGlnAsnAlaPheHisPheSerLeuSerLysAlaLeuPheAlaThr 65
 QY 770 AATTCTTATCTGA-----GACACTGTTTTTTTTTAAACAGTTAGT 814
 DB 66 LeuPheArgGlnGlnLeuSerGlyLeuThrProAlaIleValSerSerTyrGlnValSer 85
 QY 815 ACTGATGTCACACAGCAATAATTCTGATCAGATAGTCCCTGCAACAGTACGAATGT 874
 DB 86 GlnAspGlyArgPheTyrArgPheCysIleArgLysAsp----- 98

```

Qy      875 GGTTTCATAAGGTGGGAAGAAACAGCATTTTAA-----GTAACCTTTTGC  921
Db      99 -----AlaLysTrpSerAspGlySerLeuLeuLeuAlaGlnAspValIleAlaTrp  116
Qy      922 GAG-----ACGAT  930
Db      117 GlnHisThrLysGlnAlaGlyArgTyrSerLeuLeuPheGluLysLeuSerPheArgAla  136
Qy      931 TTGGTAATATATAAAGCTGTGCTGCCCTTAAGAAAAAAACCTTCCACCTT-----  984
Db      137 SerSerSerSerGluIleLeuIleGluLeuLysGlnProGlnProGlnLeuAlaIle  156
Qy      985 -----TACTGTGTCACTTATATCCCTTAAGTCCAAAGTAATATCTTATT  103
Db      157 LeuAlaSerProPhePheAlaValTyrArgProGlnLysPro-----Phe  171
Qy      1033 CTGGATATTTCCT---TTTATACCAAGAGCCTTATCAGC-----CAGTTCAG  107
Db      172 LeuSerSerGlyProPheMetProLysTrpTyrValGlnGlyHisThrLeuValLeuGln  191
Qy      1078 AACCAACCAATATACGACCAACCATACCAACCAATACCAACCAATATAGCTTAAATATC  113
Db      192 LysAsnProTyrTyrTyrAspHisAlaHisValGlnLeuHisSerIleAspPheArgIle  211
Qy      1138 ATCAGTAAGTGGCAGCA-----TATATTTCAGC-----CTTGATTGCCAAATA  118
Db      212 IleProAsnIleTyrThrAlaLeuHisLeuLeuArgArgGlyAspValAspTrpValGly  231
Qy      1183 GCACCTTCTACAT-----CCATATCTGCATCTTTCACATATACACATACATACCA  123
Db      232 GlnProTyrHisGlnGlyIleProPheGlu-LeuArgThrTrpSerAlaLeuTyrThrHis  251
Qy      1234 CTTTGTGTAAGTGCATCTAAGAGCAATGCGAAATGTAAACCTATATATTACTGATACT  129
Db      251 StyProValAspGly-----ThPheTrpLeuIle  261
Qy      1294 CTTTGGTTCCAGA  1306
Db      261 eLysAsnProLys  265

RESULT 10
US-09-971-228-7
: Sequence 7, Application US/09971228
: Patent No. US2002015512A1
: GENERAL INFORMATION:
: APPLICANT: Liao, X. Charlene
: APPLICANT: Masuda, Esteban
: APPLICANT: Chu, Peter
: APPLICANT: Pardo, Jorge
: APPLICANT: Li, Congfen
: APPLICANT: Zhao, Haoran
: APPLICANT: Jiang, Yingping
: APPLICANT: Rigel Pharmaceuticals, Incorporated
: TITLE OF INVENTION: EGS: Modulators of Lymphocyte Activation and Migration
: FILE REFERENCE: 021044-000310US
: CURRENT APPLICATION NUMBER: US/09/971,228
: PRIOR FILING DATE: 2002-04-19
: PRIOR APPLICATION NUMBER: US 60/284,763
: NUMBER OF SEQ ID NOS: 15
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 7
: LENGTH: 378
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: human endothelial differentiation G-protein
: OTHER INFORMATION: coupled receptor (GPCR) 3 (EDG3)
: US-09-971-228-7

```

Score:	84.00	Matches:	41
Percent Similarity:	42.95%	Conservative:	42.95%
Best Local Similarity:	20.92%	Mismatches:	69
Query Match:	3.62%	Indels:	4
DB:	9	Gaps:	9

05-09-830-244B-2 (1-1331) x 05-09-971-228-7 (1-378),

```

697 503 CAGCGAGGACGAGGAGAGTATGATAGCAACATCTCCATTTCTTTGTT
698 64 ILETPLYSASINLUSPHEINISASARGMETGYPHEIIEGLYSANLEUALAU 83
699 698 -----TGGTTTAAACAGTAGTACT
700 84 CYSASPLEULAUAGLGLIIEALATYGLVALASINLLEUWETSERGLYLSYTHR 103
701 740 TTATCATTTATACATCTTATAGAGCCCTAAT-----TTCTATCTGAAACAGCTGT 793
702 104 PHEISER-----SERPROTHVALTPRHEULNARGINGLYSERMET 118
703 794 TTTTTTTAAACAGTAGTACT
704 119 PHEVALALEUGLYALASERTHYCYSSEULEUALALEALALEGLIADRGNLSIU 138
705 836 TTTTGATAGATAGTCCCTGTCACAGTAGCAAA----- 871
706 139 THMETIELYUWENARGPROGLYASRALASENLYARGNISARGVALRHEULEULIE 158
707 872 -----TGGGTGTTCTATAAGTGGGAAGAAACAGCATTTTAAAGTAACCTTTTGGGAGA 925
708 159 GLYMETCYSRPLEULALEPHEIIEGLIUALAU-PROILEUGLYTPRASN 178
709 926 CT-----GATTGAGATATAATAACCTGTGCTCCCTTAAAGAAAACCCCTTC 979
710 178 YLEUHNISANLEULROASPCYSSERTIIELEUPROEULYUYSERGLYS---TYTI 197
711 980 ACCTTACGTCCT---ATTATATCCCTTAGTTCGCAAGTATATATTTCTGG 1034
712 197 LEALPHECYSLLESERTIIEPHEIIEALALEULVALTHIRLEVALIIEULYGLAIA 217
713 1037 AATATGCTTTTATACCAAGACCTTATACAGCAGTCCAGAACACACAGTATAGCACA 1094
714 217 TGLIETYPHEULVALYSSERISERARGLYVALALASN-----HLSA 233
715 1097 ACCATACCAACACAAATACCAACAATATAGCTTTTATATATCATC 1140
716 233 SNASNSERGIDARGSERMETALALEULNARGTHIRVALIIE 247
717
718 RESULT 11
719 US-09-842-316-7
720 : Sequence 7, Application US/09842316
721 : Patent No. US20020099191A1
722 : GENERAL INFORMATION:
723 : APPLICANT: KOSTENIS, Eva
724 : APPLICANT: GASSENHUBER, Johann
725 : TITLE OF INVENTION: EDG8 RECEPTOR, ITS PREPARATION AND USE
726 : FILE REFERENCE: 38005-147
727 : CURRENT APPLICATION NUMBER: US/09/842.316
728 : CURRENT FILING DATE: 2001-04-26
729 : PRIOR APPLICATION NUMBER: EP 116589.3
730 : PRIOR FILING DATE: 2000-08-01
731 : PRIOR APPLICATION NUMBER: EP 108858.2
732 : NUMBER OF SEQ ID NOS: 9
733 : SOFTWARE: PatentIn version 3.1
734 : SEQ ID NO 7
735 : LENGTH: 378
736 : TYPE: PRT
737 : ORGANISM: Homo sapiens
738 US-09-842-316-7

```

Score: 84.00
Percent Similarity: 42.35%
Best Local Similarity: 20.92%
Query Match: 10
Matches: 41
Conservative: 42
Mismatch: 69
Indels: 44
Gaps: 9

US-09-830-244b-2 (1-1331) x US-09-842-316-7 (1-378)

```

OY 653 CTCGAGACACAGACAGACAGTGTAGACACTCCATCTCTTGT----- 697
DB 64 ILETPLYSASnslsnyphenhsasnaargmetlyrphenlelglalsnleuAlaleu 83
OY 698 -----TTGTTTTTAAGACAGACACTGTCTCAAAAAAGACAT 739
DB 84 CysAspLeuAlaIgllyleAlatylrLysValasnlleuemetsercilylsysthr 103
OY 740 TTATCATTTAAACATCTTTAGAGCCCTAAT-----TTCTTATCTGAAGCAGCTGT 793
DB 104 PheSerLeu-----serProthrValtrPheLeuArgIgllyserMet 118
OY 794 TTTTCTTTTAAACAGTGTAGTACT-----GATGTCAACAGACAATA 835
DB 119 PheValAlaLeuGlyAlaSerThrcysSerLeuAlaIleAlaIleGluArgHisLeu 138
OY 836 TTTCTGATCAGATAGTCCCTGTCAACAGTACAA----- 871
DB 139 ThrMetIleLysMetArgProtyrAspAlasnlSargHisArgValPheLeuLeuIle 158
OY 872 -----TGCGTTTCATTAAGTGGAGAGAAACGATTTAAAGTACTTTTGGAGA 925
DB 159 GlyMetCysTrpLeuIleAlaPheThrLeuGlyAlaLeu--ProIleLeuGlyTrpSnc 178
OY 926 CT-----GATTGATTAATATAAAGCTGTCTGTCTTAAAGAAAAAACCTTCC 979
DB 178 YsLeuHisAsnLeuProAspCysSerThrIleLeuProLeuTyrsLysLys--TyrI 197
OY 980 ACCTTACTGTGC---ATTATATCCCTTACTTCCAAAGTAAATATCTTATTTCTGG 1036
DB 197 leAlaPheCysIleSerIlePheThrAlaIleLeuValThrIleValIleLeuTyAla 217
OY 1037 ATATGCTTTTATACCAAGAGCTTATACAGCCAGTCCAGAACCAACACTATACGACA 1096
DB 217 rgIleTyPheLeuValLysSerSerArgLysValAlaasn-----HisA 233
OY 1097 ACCATACCAACCAATACCAACAATATACGTTTATATATCATC 1140
DB 233 snasnsrGluArgSerMetAlaLeuLeuArgThrValValIle 247

```

RESULT 12
US-09-731-030A-18
Sequence 18, Application US/09731030A
Patent No. US20020142375A1
GENERAL INFORMATION:
APPLICANT: MUNROE, Donald G
APPLICANT: GUPTA, Ashwani K.
APPLICANT: ZASTAMNY, Roman L.
TITLE OF INVENTION: MAMMALIAN EDG-7 RECEPTOR HOMOLOGS
FILE REFERENCE: 8074-0015
CURRENT APPLICATION NUMBER: US/09/731,030A
CURRENT FILING DATE: 1998-12-29
PRIOR APPLICATION NUMBER: 60/070,184
NUMBER OF SEQ ID NOS: 21
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 18
LENGTH: 378
TYPE: PRT
ORGANISM: Unknown
FEATURE:
OTHER INFORMATION: Description of Unknown Organism: Mammalian EDG
US-09-731-030A-18

Alignment Scores:

Pred. No.: 1.37
Score: 84.00
Percent Similarity: 42.35%
Best Local Similarity: 20.92%
Query Match: 10
Matches: 378
Conservative: 41
Mismatch: 69
Indels: 44
Gaps: 9

US-09-830-244b-2 (1-1331) x US-09-731-030A-18 (1-378)

```

OY 653 CTCGAGACACAGACAGACAGTGTAGACACTCCATCTCTTGT----- 697
DB 64 ILETPLYSASnslsnyphenhsasnaargmetlyrphenlelglalsnleuAlaleu 83
OY 698 -----TTGTTTTTAAGACAGACACTGTCTCAAAAAAGACAT 739
DB 84 CysAspLeuAlaIgllyleAlatylrLysValasnlleuemetsercilylsysthr 103
OY 740 TTATCATTTAAACATCTTTAGAGCCCTAAT-----TTCTTATCTGAAGCAGCTGT 793
DB 104 PheSerLeu-----serProthrValtrPheLeuArgIgllyserMet 118
OY 794 TTTTCTTTTAAACAGTGTAGTACT-----GATGTCAACAGACAATA 835
DB 119 PheValAlaLeuGlyAlaSerThrcysSerLeuAlaIleAlaIleGluArgHisLeu 138
OY 836 TTTCTGATCAGATAGTCCCTGTCAACAGTACAA----- 871
DB 139 ThrMetIleLysMetArgProtyrAspAlasnlSargHisArgValPheLeuLeuIle 158
OY 872 -----TGCGTTTCATTAAGTGGAGAGAAACGATTTAAAGTACTTTTGGAGA 925
DB 159 GlyMetCysTrpLeuIleAlaPheThrLeuGlyAlaLeu--ProIleLeuGlyTrpSnc 178
OY 926 CT-----GATTGATTAATATAAAGCTGTCTGTCTTAAAGAAAAAACCTTCC 979
DB 178 YsLeuHisAsnLeuProAspCysSerThrIleLeuProLeuTyrsLysLys--TyrI 197
OY 980 ACCTTACTGTGC---ATTATATCCCTTACTTCCAAAGTAAATATCTTATTTCTGG 1036
DB 197 leAlaPheCysIleSerIlePheThrAlaIleLeuValThrIleValIleLeuTyAla 217
OY 1037 ATATGCTTTTATACCAAGAGCTTATACAGCCAGTCCAGAACCAACACTATACGACA 1096
DB 217 rgIleTyPheLeuValLysSerSerArgLysValAlaasn-----HisA 233
OY 1097 ACCATACCAACCAATACCAACAATATACGTTTATATATCATC 1140
DB 233 snasnsrGluArgSerMetAlaLeuLeuArgThrValValIle 247

```

RESULT 13
US-10-037-616-21
Sequence 21, Application US/10037616
Patent No. US20020123148A1
GENERAL INFORMATION:
APPLICANT: English, Denis
APPLICANT: Kovacs, Richard J.
APPLICANT: Rizzo, Maria T.
APPLICANT: Silva, Daniel T.
TITLE OF INVENTION: Sphingolipid Compositions and Methods for Their Therapeutic Use
FILE REFERENCE: 7042-119
CURRENT APPLICATION NUMBER: US/10/037,616
CURRENT FILING DATE: 2002-04-19
PRIOR APPLICATION NUMBER: US 60/243,887
NUMBER OF SEQ ID NOS: 27
SOFTWARE: PatentIn version 3.1
SEQ ID NO 21
LENGTH: 378
TYPE: PRT
ORGANISM: Homo sapiens
US-10-037-616-21

Alignment Scores:

Pred. No.:	1.37	Length:	378
Score:	84.00	Matches:	41
Percent Similarity:	42.35%	Conservative:	42
Best Local Similarity:	20.92%	Mismatches:	69
Query Match:	3.62%	Indels:	44
DB:	12	Gaps:	9

US-09-830-244B-2 (1-1331) X US-10-037-616-21 (1-378,

```

935 CACGCGGAGCAGGAGCAGTATAGTAAACCTCCATCTCTTTGGT
      :::::::::::
64 ILETPLYSAsnAsnLysPheHisAsnArgMetLysPhePheIleGlyAsnLeuAlaLeu
      :::::::::::
698 -----TGTCTTTTAAAGACAGACACTGTCTCCAAAAGAGACAT
      :::::::::::
84 CysAspLeuLeuAlaGlyIleAlaTyrLysValAsnIleLeuMetSerGlyLysThr
      :::::::::::
740 TTATCATATATACATCTTATTAGAGCCCAAT-----TCTTTATGTAAGCAGCTGT
      |||||
104 PheSerLeu-----SerProThrValTyrPheLeuArgGlyGlySerMet
      :::::::::::
794 TTTTCTTTTAAACAGTTAAGTACT-----GAGTGCACAGACAAATA
      |||
119 PheValAlaLeuGlyAlaSerThrCysSerLeuLeuAlaIleAlaIleGlyAlaThrHisLeu
      |||||
836 TTTCTGATGACATAGTCCCTGTCACACAGTACCAAA-----
      :::::::::::
139 ThrMetIleLysMetArgProTyrAspAlaAsnLysArgHisArgValPheLeuLeuIle
      :::::::::::
872 -----TGTGAGTTTCATTAAGTGGGAAAGAAACACCATTTTAAGTAAGTCTTTGGGAGA
      |||
159 GlyMetCysTrpPheLeuIleAlaPheThrLeuGlyAlaLeu--ProIleLeuGlyTyrPasnC
      :::::::::::
926 CT-----GATTGAGTAATTAATTAACACTGTGCTCCCTTAAGAAAAAACCCTCC
      :::::::::::
178 LysLeuHisAsnLeuProAspCysSerThrIleLeuProLeuTyrSerLysLys--TyrI
      :::::::::::
980 ACCTTTACTGTGTC---ATTTATATCCCTTAGTCTCCAAAGTAATATATCTTATTTCTGG
      :::::::::::
197 LeuAlaPheCysIleSerIlePheThrAlaIleLeuValThrIleValIleLeuTyrAla
      :::::::::::
1037 ATATGCTTTTATATCAACAGCCTTATACAGCACTTCAGACACACCACTATACGCACA
      |||
217 rGlyLeuTrpPheLeuValLysSerSerSerArgLysValAlaAsp-----HisA
      :::::::::::
1097 ACCATATCAACAACAATACCAACATATACGCTTTTAATATATCAGC
      ||:::::
233 snAsnSerGluArgSerMetAlaLeuLeuAlaGlyThrValValIle

```

RESULT 14
US-09-876-790-4
: Sequence 4, Application US/09876790
: Publication No. US20030091532A1
: GENERAL INFORMATION:

US-09-876-790-4	
Alignment Scores:	
Pred. No.:	1.69
Score:	64.00
Percent Similarity:	33.66%
Best Local Similarity:	23.41%
Query Match:	3.63%
DB:	9
	Gaps: 12
	Length: 65
	Matches: 48
	Conservative: 20
	Mismatches: 50
	Indels: 86

US-09-830-244B-2 (1-1331) X US-09-876-790-4 (1-696)

[illegible]

RESULT 15
US-10-011-548-35
; Sequence 35; Application US/10011548
; Publication No.: US20030055218A1
; GENERAL INFORMATION:
APPLICANT: Timans, Jacqueline C.

Debets, Johannes Eduard Maria
Antonius
Sana, Theodore R.
Bazan, J. Fernando
Kastlein, Robert A.

TITLE OF INVENTION: Human Receptor Proteins; Related Reagents and Methods

NUMBER OF CLAIMS:

NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: DMAX Research Institute
STREET: 901 California Avenue

QY 1155 TCTCTGCAGTACGATGATATATTAACGTATATGTTGGATATGCTGGATGCTT 1096
Db 3 CysThrAspTTPSerIleAspIleLysLysTyr-----GlnValLeuValGlyGlnPro 20
QY 1095 GGGCGTATAGTGGTTGCTG-----GAACTGGCTAAT 1063
Db 21 ValAlaGlyLysCysAlaLeuPheTyrGlyTyrIleArgThrAsnTyrSerLeuAlaGln 40
QY 1062 AAGCCT-----CTTGGTATATAAAGCAATATCCAGAAATAGATATTAACCTT 1015
Db 41 SerAlaGlyLeuSerLeuMetTTPyrLysSerSer----- 52
QY 1014 TGGAACTAAGGGGATATATAATGACACAGTAAGAAGGTGTTTCTTCTTAAGGA 955
Db 53 ---GlyProGlyAspPheGlnGluProIleAlaPheAspGlySerArgMetSerLysGln 71
QY 954 GACCAAGATTTTATTATTACTCAAAATCAGTCCCAAAAGTACTTTAAATATGCTGTT 895
Db 72 GluAspSerIle----- 75

Search completed: June 24, 2003, 19:27:51
Job time : 57.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 24, 2003, 19:16:52 ; Search time 44 Seconds

(without alignments)
5816.133 Million cell updates/sec

Title: US-09-830-244b-2

Sequence: 1 ccatgttttagatcaaa.....gccttccaatgtcacttg 1331

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-O=/cgn2.1/USPRO-SPool/US0983024/unal_24062003_130002_10054/app_query.fasta.1.1479
-DB=PIR.73 -QFMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blousum62 -TRANS=human40.cdi -LIST=45
-DOCALL=200 -THR_SCORE=pcpt -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORMEXT=HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0983024 -CGN.1.1.71 -runat_24062003_130002_10054 -NCPU=6 -ICPU=3
-NO_MMAP -LARGOUDRY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

PIR.73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	136	5.9	62	1	SBHUP
2	91.5	4.0	949	2	P90086
3	91.5	4.0	1138	2	S64484
4	91	3.9	2708	2	T03079
5	91	3.9	2819	2	T09080
6	90	3.9	691	2	T46476
7	89	3.8	445	2	A44848
8	88.5	3.8	474	2	T31104
9	88	3.8	1005	2	P90099
10	87.5	3.8	332	2	T12023
11	87	3.8	855	2	T17460
12	86	3.7	2028	2	T08025
13	85.5	3.7	360	2	A53611
14	85.5	3.7	1070	2	G84982

15	85	3.7	61	1	SBHUP	stathurin precursor
16	85	3.7	567	2	S58750	NADH2 dehydrogenas
17	85	3.7	577	2	T04229	ABC-type transport
18	85	3.7	580	2	T18439	hypothetical prote
19	85	3.7	600	2	T18446	hypothetical prote
20	85	3.7	620	2	A58932	cytochrome C-type
21	85	3.7	2391	2	T18410	cardamoyl-phosphat
22	84	3.6	378	2	JC5245	G protein-coupled
23	84	3.6	651	2	F90111	DNA primase [impor
24	84	3.6	748	2	S54505	hypothetical prote
25	83.5	3.6	331	2	T26437	ATP-dependent RNA
26	83.5	3.6	515	2	P90366	tubulin beta chain
27	83	3.6	442	1	UBHUP	nuclear pore prote
28	83	3.6	1037	2	S73679	protein g377 - mal
29	83	3.6	3119	2	T37814	ERD1 protein - yen
30	82.5	3.6	362	2	T18414	hypothetical prote
31	82.5	3.6	475	2	S69698	replication factor
32	82.5	3.6	475	2	G59105	hypothetical prote
33	82	3.5	1847	2	E64477	interleukin-8 rece
34	82	3.5	339	2	D90106	hypothetical prote
35	81.5	3.5	386	2	S42096	lipoprotein [impor
36	81	3.5	383	2	E82897	conserved hypotec
37	81	3.5	368	2	C90558	hypothetical prote
38	80.5	3.5	835	2	F64246	beta-tubulin 1 - a
39	80	3.4	445	2	H90090	Mg148 homolog VXP
40	80	3.4	445	2	J00422	DNA-directed RNA p
41	80	3.4	960	2	S72284	hypothetical prote
42	80	3.4	1186	2	S70430	hypothetical prote
43	79.5	3.4	365	2	E90099	hypothetical prote
44	79.5	3.4	376	2	T24368	hypothetical prote
45	79.5	3.4	446	2	H90094	hypothetical prote

ALIGNMENTS

RESULT 1

SBHUP stathurin precursor [validated] - human

C:Species: Homo sapiens (man)

C>Date: 24-Apr-1984 #sequence_revision 30-Jun-1993 #text_change 08-Dec-2000

C:Accession: JH0153; A27308; B27485; A03288; A32524

R:Sablatni, L.M.; He, T.Z.; Azen, E.A.

A:Title: Structure and sequence determination of the gene encoding human salivary sea

A:Reference number: JH0153; MUID:90323623; PMID:2373365

A:Molecule type: DNA

A:Residues: 1-62 <SA2>

A:Cross-references: GB:M31077

R:Sablatni, L.M.; Carlock, L.R.; Johnson, G.W.; Azen, E.A.

A:Title: cDNA Cloning and chromosomal localization (4q11-13) of a gene for stathurin,

A:Reference number: A27308; MUID:88074310; PMID:3502720

A:Accession: A27308

A:Molecule type: mRNA

A:Residues: 1-62 <SAB>

A:Cross-references: GB:M32639; NID:g338504; PIDN:AAA60593.1; PID:g338506

R:DiChikinson, D.P.; Ridall, A.L.; Levine, M.J.

A:Title: Human submandibular gland stathurin and basic histidine-rich peptide are enc

A:Reference number: A27489; MUID:88106506; PMID:3426601

A:Accession: B27489

A:Molecule type: mRNA

A:Residues: 1-62 <DIC>

A:Cross-references: GB:M18371; NID:g338610; PIDN:AAA60600.1; PID:g338611

R:Schlesinger, D.H.; Hay, D.I.

A:Title: Complete covalent structure of stathurin, a tyrosine-rich acidic peptide whi

A:Accession: A03288

A:Molecule type: protein

A:Residues: 20-62 <SCH>

R:Oppenheim, F.G.; Hay, D.I.; Smith, D.J.; Offner, G.D.; Troxler, R.F.

C:Species: Physarum polycephalum
C>Date: 31-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 23-Feb-1997
C:Accession: A44848
J:Gen. Microbiol. 138, 229-238, 1992
A:Title: Preferential expression of one beta-tubulin gene during flagellate development
A:Reference number: A44848; MID:92211323; PMID:1556551
A:Contents: amoebae, flagellates
A:Accession: A44848
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-445 <PAU>
A:Note: sequence extracted from NCBI backbone (NCBIN:92291, NCBI:92292)
C:Superfamily: tubulin
C:Keywords: nucleus

Alignment Scores:
Pred. No.: 0.761 Length: 445
Score: 89.00 Matches: 44
Percent Similarity: 39.81% Conservative: 38
Best Local Similarity: 21.36% Mismatches: 76
Query Match: 3.83% Indels: 48
DB: 2 Gaps: 8

US-09-830-244B-2 (1-1331) x A44848 (1-445)

OY 386 TTTTCAGATCTTTGATGTCACATAATGTAATGCGCTTATTTGTAATAC 445
DB 159 TyrProThrTyrlgysleuasnHslleuValSerAlaVal-----MetSer 234
OY 446 AGCATAGTAAATAAAGTACGTAAGTAAATCTACAC-----481
DB 179 ValValGluProTyrlgysleuasnHslleuValSerAlaValGluAsnAlaAspGlu 198
OY 482 -----TTGATTTGCTGACCTTACCAATAGCCTTTGAATGACGAAGTGA 532
DB 199 ValMetCysIleAspAsnGluAlaLeuTyrlgysleuasnHslleuValSerAlaVal 218
OY 533 ACAGAGAAGAAGCATCTGTCAGAAAGAGATAGTAAATTTTGGTACTTATCTGA 592
DB 219 ThrProThrTyrlgysleuasnHslleuValSerAlaVal-----MetSer 234
OY 593 AATCAAGATGCTGC-----TTCCTGAGGCTTTTCCCTTACGATCC-----640
DB 235 GlyIleThrCysCysleuArgPheProGlyGlnLeuAsnSerAspLeuArgIleuAla 254
OY 641 ----TCAATTGAATCCCTCTGGGAGACAGACAGTATAGAACCTTCATTTCTTGT 697
DB 255 ValAsnLeuIleProPhe-----ProArgLeuHslPheLeu 267
OY 698 TTGTTTAAAGACAGACTCTGTCTCAAAAAAGACATTTATCATTA-----748
DB 268 ValGlyPheAlaProLeuThrSerArgIleValGlyTyrArgSerLeuThrValPro 287
OY 749 -----TAACTCTTATTAGAGCCCTAAT-----772
DB 288 GluLeuThrGlnGluMetPheAspAlaIleValSerAspPheArgHis 307
OY 773 -----TCTTATCTGAGAGCAGCTCTTTTAAACAGTAAAGTACGATGTCAC 826
DB 308 GlyArgTyrlgysleuasnHslleuValSerAlaValGlyTyrArgSerLeuThrValPro 327
OY 827 AGACAAAATTTCTGATAGATAGTCCCTGTCACACATAGCAAAATGCTTTGTAAG 886
DB 328 GluGlnMetLeuAsnValGln-----AsnIleValSerAlaValGlu 343
OY 887 TGGAGAGAAACAGCATT 904
DB 344 TrpIleProAsnAsnIle 349

RESULT 8
T31104
hemolysin accessory protein homolog - Haemophilus ducreyi (fragment)

C:Species: Haemophilus ducreyi
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T31104
J:Gen. Microbiol. 180, 6013-6022, 1998
A:Title: Haemophilus ducreyi secretes a filamentous hemagglutinin-like protein.
A:Reference number: 220984; MID:99030326; PMID:9811662
A:Accession: T31104
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-474 <MAR>
A:Cross-references: EMBL:AF057696; MID:93929021; PID:93929022; PIDN:AACT9760.1
A:Genetics:
A:Gene: lpsB

Alignment Scores:
Pred. No.: 0.867 Length: 474
Score: 88.50 Matches: 42
Percent Similarity: 38.46% Conservative: 33
Best Local Similarity: 21.54% Mismatches: 71
Query Match: 3.83% Indels: 49
DB: 2 Gaps: 10

US-09-830-244B-2 (1-1331) x T31104 (1-474)

OY 885 TTTATGAACACATTTGCTACTGTTGACAGGGGAC---TATCTGATCAGAATTTTGT 829
DB 82 PheValIleProAsn-----LeuSerGlyIlePheSerIleHslAspLeuAsp 98
OY 828 CTGTGACATCAGTACTTAATCTGTTTAAAAAACA-----GTGCTTTCAGAT 778
DB 99 GluLeuValGluValLeuAsnThrValAsnIleValAsnIleValAsnIleValAsnIle 118
OY 777 AAGAAATAGAGGGCTCTAATAAGATGTTATATGTAATGCTTTTGGAGACA 718
DB 119 LysAlaIleGlyLys-----SerAsn 125
OY 717 GTCTGTGCTTAAAAAACAAGAAATGAGAGTTCATCTGCTGCTCC 658
DB 126 LeuAsnIleLeuThrGlnArgIleValAsnIleValAsnIleValAsnIleValAsnIle 144
OY 657 CAGAGGGATTCATAGAGATCGTACAGAGAAACACCTGCAGGGAGACAGATCTT 598
DB 145 AsnSerGly-----LysGlnAsnAsnGlnAsnGlnAsnGlnAsnGlnMet 158
OY 597 GGATTCAGATAAGTACCAAAAAATATATGCTATCTTTCGACAGATGCTTTTC 538
DB 159 ThrLeuAsnValSerIlePheAspLeuGlyThrAsn-----171
OY 537 TCTGTAACTTTCACCTTACCTATTCAAAAAGCTATTGGTAGAGTGCAGCAATCCAGTGT 478
DB 172 -----AspValIlePheSerPheIleThrGlyTyr-----ArgLeuIleValGluThrIle 187
OY 477 AGATTTCCTTAACAGATTTATTAATGCTATATCCACATTAAGCCATTATCAATA 418
DB 188 LysAsnThrGlnIleAsnIleValSerLeuSerIleGlnProPheSerIleTyrlleu 207
OY 417 ATGCAATTAAGTACATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 376
DB 208 -----GluIleLysIleSerGlnSerAlaIleValAsnIleValGluLeuThrGlyPhe 223
OY 375 ----ACATTTCCCAACAAGAAAGAAATCAATGCTCTCTATTA 334
DB 224 TyrThrTyrlgysleuasnHslleuValSerAlaValGlyTyrArgSerLeuThrValPro 343

RESULT 9
F90099
hypothetical protein orf1005 (imported) - Guillardia theta nucleomorph
C:Species: nucleomorph Guillardia theta
A:Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C:Accession: F90099
R:Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.;

Nature 410, 1091-1096, 2001
 A:Title: The highly reduced genome of an enslaved algal nucleus.
 A:Reference number: A99082; MUID:11323671; PMID:11323671
 A:Accession: F90099
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1005 <DOD>
 A:Cross-references: GB:AF165818; NID:g13794550; PIDN:AAK39925.1; GSPDB:GN00150
 C:Genetics:
 A:Gene: orf1005
 A:Map position: 1
 A:Genome: nucleomorph
 C:Keywords: nucleomorph

Alignment Scores:
 Pred. No.: 1.07 Length: 1005
 Score: 88.00 Matches: 99
 Percent Similarity: 37.21% Conservative: 45
 Best Local Similarity: 25.38 Mismatches: 125
 Query Match: 3.81% Indels: 118
 Gaps: 22

US-09-830-244B-2 (1-1331) x F90099 (1-1005)

```

OY 999 AATAATGACACGATGAAGGTGAGAGGTTTCTTAAGGAGACAGAGTTTAT 940
Db 132 IleaAsnThrValSerLeuLysAsnLeuPheTy--LysCysAspAsnLeuPheIle 150
OY 939 ATTACTCAA-----
Db 151 LeuIleGluLysTySerGlyCysArgThrLeuLysLeuPheLysAsnLysLeuArgLeu 170
OY 930 -----ATCAGTCTCCCAAAAAGTACTTAAATGCTG 898
Db 171 LeuGlnPheAsnLeuLysAsnAsnThrIleAsnThrLysLysLysTy--LysSerIle 189
OY 897 TTTTCT-----TCCCACTTATGAACACCACTTGTCTAGCTGTGACAGG 853
Db 190 PheSerLysLysLysPheLysPheGlnLysGlnLysLysLysLeuLeu----- 206
OY 852 GACTATCTGATCAGAAATATTGTCTGTGACATCAGTACTTAACTTTAAAAA 793
Db 207 AspArgTyrlleArg-----LeuPheSerIleLeuAspAsnTyrls----- 220
OY 792 ACACTGCCCTCAGTAAGAAATTTGGGCTCTAATAAGATTATATGATAATGCTCT 733
Db 221 -----IleAspTyrlsLeuPhePheTyrllePheHisTyrlTpgInThrLys 237
OY 732 TTTTGTGAGACAGAGTCTGTCTTAAACAAAACAAAGAAATGGAGAGTTCTACTA 673
Db 238 IleLeu-----LysLysAsnIleArgGluLysLysAspCysAsnPheLysValIleLeu 255
OY 672 ACTGTCTCTGCTCCAGAG-----
Db 256 LysIleAsnLeuPheThrLysPheGluLysLysAsnAsnTyrlSerAsnTyrlSerThrIle 275
OY 651 -----GGATTCAATGAGATCTGTAAGAGAAAACACCTGCAGGGAGACATCTT 598
Db 276 PheTyrlPhePheAsnSerAsnLysLysIleAsnAsnIleIleLysAsnLeuGlnPhe 295
OY 597 GGATTTCAGATAAGTACCAAAAATATTAGTATCTTTCTGGCAGACATCCCTTTTC 538
Db 296 AsnPhe--AspLysLeu-----AsnSerLysTyrllePhe----- 306
OY 537 TCTGTTAACACTTTCAGTACTTCAAAAGCTATTGTGTAGAGGTGCAGCAATCCAAGTGT 478
Db 307 -----SerLysSerLeuAspLeuArgProLysAsn----- 316
OY 477 AGATTACTGTACAGATTATTAATGCTGTATGCTGATTCACCAATAAGGCATATTACATA 418
Db 317 -----GluSerIleMetAsnMetIlePheValThrGluSerAsnIleLysAsnA 333
OY 417 ATGGCATTTAGTGACATCAAAAAGATCTGAAAATGCTAATGGACATTTCCCAAAAAGA 358
  
```

```

Db 333 nValAlaValLysArgIleLysValLys-----LysLysSerI 347
OY 357 AAGAAATCCAAATGGCTCTTAAAGGAGAAATATACCAATGCAGACAGATGAATTT 298
Db 347 yslleuIleTyrlsIleSerLysGlnIleSerIleThrThrPheLysTyrlPheLysIleI 367
OY 297 CTGTCAAAAGAGTACTTCTCAACAGCAATCAGCAAAATGTTCCCAATTAAGCCCTGG 238
Db 367 lePheLysLysLysTyrlPhe-----GlnAsnLysIleLeuLeuSerLeuPheIleI 384
OY 237 GGAGCCACATTAGC---CACTATCTCAGATACTGATTATTGTTT-----AGCCT 190
Db 384 leuMetIleSerLysHisTyrlle--AspLys-IleValIleTyrlLysLysAsnSerIle 403
OY 189 TAGGGGCTCCCAAT-----ATTCGCCCTCAT-----CTGCTAACTTCAACTT 148
Db 403 uLeuGlyTyrlAsnIleIleIleAsnGlnIleGlyValArgGlyValIleuAsn--AsnIle 422
OY 147 GTTTACTCAACAGTCTGATTAATTCAGCCAGTACTT-----ACATTC 100
Db 422 uSerLysProLeuAspSerLysTyrlArgAsnSerAsnIleTyrlAspLysCysHisIleVal 442
OY 99 CTATGGCAGTGCAGAAATTCACAGCATTTGTAACACTGTAATTTACGTTTGTATCAG 40
Db 442 lPhePheSerArgLysTyrlPheThrIlePheLys-----AsnTyrlPhePhe----- 457
OY 39 AGACAGCTTACCG 27
Db 458 ----SerLeuPro 460
  
```

RESULT 10
 T25023
 Hypothetical protein T20B3.5 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
 C:Accession: T25023
 R:Moritome, B
 submitted to the EMBL Data Library, November 1996
 A:Reference number: Z19970
 A:Accession: T25023
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-332 <MWL>
 A:Cross-references: EMBL:Z81593; PIDN:CA804743.1; GSPDB:GN00023; CESP:T20B3.5
 A:Experimental source: clone T20B3
 A:Gene: CESP:T20B3.5
 A:Map position: 5
 A:Introns: 110/1; 246/3
 C:Superfamily: Caenorhabditis hypothetical protein C4967.2

Alignment Scores:
 Pred. No.: 1.07 Length: 332
 Score: 87.50 Matches: 64
 Percent Similarity: 33.99% Conservative: 40
 Best Local Similarity: 20.92% Mismatches: 87
 Query Match: 3.77% Indels: 115
 Gaps: 2

US-09-830-244B-2 (1-1331) x T25023 (1-332)

```

OY 529 GTTAACAGAGAAAGAGCATGCTGCAGAAAGAGATAGCTAATTTTGTG----- 579
Db 89 ValLeuLysGluIleGlyValSerValGluAlaGlnVal--TyrlPheValIleuVal 107
OY 580 -----
Db 108 GlyAlaTyrlMetMetValAlaIleSerIleValPheGluAsnArgLeuPheValIleuVal 127
OY 589 CTGAAATCCAAAGTGTG-----CTTCCCTGCAGGTGTT----- 624
Db 128 MetThrAsnLysMetLeuHisLysPheAlaLeuProIleTyrlIleHisTyrllePhe 147
  
```

QY 625 -----TTCCTGTCAGATCCTCATTTGAATCCCTCTGGAGCAGACAGAGTAGTAGT 678
 Db 148 ProThrIleValLeuProSerLeuValLysLeuPro----- 159
 QY 679 AACTGTCATCTTCTGTTGTTTGTATTTTAAAGACAGACTCTGTCTCAAAAAGAGACA 738
 Db 160 ----- 167
 QY 739 TTTATCATTAATACATCTTATAGACCCCTAATTTCTTATCTGAAGCAGCTGTTT 798
 Db 168 PheLeuLeu----- 170
 QY 799 TTTTAAACAGTAAGTACTGATGTCACAGACAAATATTTCTGATCAGATAGTCCCTGT 858
 Db 171 -----LysTyrGlyCys-----ValProProTyr 178
 QY 859 CAACAGTAGCAAAATGTGTTTCATTAAGTGGAGAGAAACAGACATTTTAAAGTAAC--- 915
 Db 179 ValAspLeuGluArgValPheTyrLeuIleIleThrLysArgTyrPheLeuIleThrCys 198
 QY 916 -----TTTGG-----GAGACTGATTTG 933
 Db 199 AlaValPheIleCysThrMetPheAlaGluValTyrPhePheAlaLeuValThrAspArg 218
 QY 934 AGTAATTAATAAACTGTGCTCCCTTAAGAAAAAAACCTTCAACCTTTAC----- 987
 Db 219 LeuLeuLysLysGlnMetThrLysThrMetSerGlnLysThrPheAspLeuHisLysLys 238
 QY 988 -----TGTGCATTTATATCCCTTACTGTCCAAAATTAATTAATCTTATTTCTGAT 1038
 Db 239 PheGlnArgAlaPheIleLeuGlnLeuLeuLeuPro---PheIleIleValPheLeuPro 257
 QY 1039 ATTCCTTTTATACCAAGACCTTATGAGCCAGTTCAGAACCAACCACTTATACGACAC 1098
 Db 258 IleSerTyrIleGlyValIleThrCysIleSerGlnLysAsnGlnPhePhe----- 274
 QY 1099 CATACCAACCAATACCAATATACGTTTAAATATCATCATGATCATGACAGACATGA 1158
 Db 275 -----AsnAsnLeuThrLeuIleIleIleSerHisGly----- 286
 QY 1159 TTATGAGGCTGATGGAATAACGACTTCTACATTCATTCATCTTTCATATACAT 1218
 Db 287 PhePheSerThrIleAlaMetIleAlaLeuHisAlaProTyrArgGlnPheThrGlnLys 306
 QY 1219 ATACACTACTACCACTTTTGTGAGA---TCATCTAAGACCAATGCGAATGAAACCC 1275
 Db 307 Ile-----PheProCysLeuLysArgPheSerSerAlaThrSerThrAlaPro 324
 QY 1276 TATATTTTACTGATACT 1293
 Db 325 GlnSerLeuMetIleThr 330
 RESULT 11
 T17460
 disease resistance protein - tomato
 C:Species: Lycopersicon esculentum (tomato)
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-May-2000
 C:Accession: T17460
 R:Parisi, M.; Jones, J.D.
 Proc. Natl. Acad. Sci. U.S.A. 96, 5850-5855, 1999
 A>Title: Recombination between diverged clusters of the tomato Cf-9 plant disease resist
 A:Reference number: Z18801; MUID:99254130; PMID:10318973
 A:Accession: T17460
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-855 <PAR>
 A:Cross-References: EMBL:AF119040; NID:94235640; PID:94235642; PIDD:AA13302.1
 A:Gene: NLOC
 Alignment Scores:
 Pred. No.: 1.34 Length: 855

Score: 87.00 Matches: 71
 Percent Similarity: 34.06% Conservative: 38
 Best Local Similarity: 22.19% Mismatches: 116
 Query Match: 3.76% Indels: 95
 Db: 2 Gaps: 15
 US-09-830-244B-2 (1-1331) x T17460 (1-855)
 QY 906 AAAATGCTGTTTCTTCCCATTTATGAAACCAATTTGCTACTGTGACAGGACATAT 847
 Db 5 LysLeuValPhePheMetArgTyrVal-----PheLeuPheIleLeuValSerSer 22
 QY 846 CTGATCAGAAATATTTGCTGTGATCATCAGTACTTAAGCTTTTAAAAAACAAGCAGT 787
 Db 23 SerLeuArgHisLeuCys----- 28
 QY 786 CCTTCAGATTAAGAAATTAAGGCGCTTAATAGATTTATAT- 745
 Db 29 ProGluAspGlnAlaLeu---AlaLeuLeuGlnPheLysAsnMetPheThrValAsnPro 47
 QY 744 -----GATTAATGCTCTTTTGTGAGACAGAGCTCTGCTTAAAAAACAACA 694
 Db 48 AsnAlaPheHisTyrCysProAspIleThrGlyArgGluIleGlnSerTyrProArgThr 67
 QY 693 AAGAAATGAGAGATTTCTACTACTGCTGCTGCTCCAGAGGGGATTCAGATGATGCT 634
 Db 68 LeuSerTyrAsn---LysSerThrSerCysSerTyrPaspGlyValHisCysAspGlu 86
 QY 633 AAGAAC-----GAAACAACCTGCAGGGGAGACAGACATCTTGA--- 595
 Db 87 ThrThrGlyGlnValIleAlaLeuAspLeuGlnLeuGlnGlyLysPheHisSerAsnSer 106
 QY 594 -----TTTCAGATTAAGTACCAAAATATTAATGATCTTTCTGACAGATGCTCT 541
 Db 107 SerLeuPheGlnLeuSerAsnLeuLysArgLeuAspLeuSerPhe----- 121
 QY 540 TTCTCTGTTAACTTTGAGTCAATTCAAAGAGCTAATGGTAGAGTGACCAATTCACAG 481
 Db 122 -----AsnAspPheThrGlySerPro-----IleSerProLysPhe 133
 QY 480 TGTAGATTTACTGTAAGAGCTTATTAATTAACATGCTGTAATTCACCAATAGGCAATTCACA 421
 Db 134 GlyIlePheSerAspLeuThrHisLeuAspLeuSerHisSerPheThrGlyLeuIle 153
 QY 420 ATTAATGCGATTAAGTGCATCAATCAAAAGATCGTGAATAATGATGACATTTCCACAAAA 361
 Db 154 ProPheGluIleSerHisLeuSerLys-LeuHisValLeuArg----- 167
 QY 360 GGAAGAATTCCAATGCTCCTATTAAGGAGATATAGCAATGCAAGCAGAAATGAAA 301
 Db 168 -----IleArgGlyGlnTyrLysLeuSerLeuValProHisAs 180
 QY 300 TTTCTGCAAAAGAAAGTACTTCTCAAGACAAATCAGGAAACTGTTCCCAATTAAGGCC 241
 Db 180 nPhe-----GluLeuLeuLysAsnLeuThrGlnLeuArg--- 192
 QY 240 TGGGAGCCATTAAGCAGTATGCAATGATGATTAATTTGTTTGAAGCTTAAGGGCCT 181
 Db 193 -----AspLeuGlnLeuGlnSerIleAsnIleSerSerThrValProse 207
 QY 180 CAATATTTCCCTTCATCTGTTAACTTCAACTTG-----AAATTCACACCAAG 115
 Db 207 rAsnPheSerSerHisLeuThrAsnLeuArgLeuProPheThrGlnLeuArgGlyIleLe 227
 QY 146 -----TTTACTCACAAGCTGAT-----AAATTCACACCAAG 115
 Db 227 uProGluArgPhePheHisLeuSerHisLeuGlnSerLeuAspLeuSerPheAsnPro-- 246
 QY 114 TAACCTTAATCTCCCTATGCGAGTGCAGAAATTCAGACATTTGTAACACATGTAAT 57
 Db 247 -GlnLeuThrValArgPheProThrThrLysTyrPaspSerSerAlaSerLeuValAsn 265
 RESULT 12

T08025

DNA-directed RNA polymerase beta' chain homolog - Chlamydomonas reinhardtii chloroplast

N:Alternate names: rpoC2 protein

C:Species: chloroplast Chlamydomonas reinhardtii

C:Date: 21-May-1999 #sequence_revision 21-May-1999 #text_change 08-Oct-1999

C:Accession: T08025

R:Mutation, S.: Purton, S.

submitted to the EMBL Data Library, May 1996

A:Description: The chloroplast rpoC2 gene of Chlamydomonas reinhardtii.

A:Accession: T08025

A:Status: preliminary; translated from GB/EMBL/DDBT

A:Molecule type: DNA

A:Residues: 1-2028 <MOR>

A:Cross-references: EMBL:U57326; NID:91354831; PIDN:AA01997.1; PID:91354832

C:Genetics:

A:Gene: rpoC2

A:Genome: chloroplast

C:Keywords: chloroplast

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

US-09-830-244b-2 (1-1331) x T08025 (1-2028)

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-360 <RES>

A:Cross-references: EMBL:U11869; NID:9511801; PIDN:AA06056.1; PID:9511803

A:Accession: I38712

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-15 <RE2>

A:Cross-references: EMBL:U11872; NID:9511808; PIDN:AA064380.1; PID:9511809; EMBL:U11876; NID:9511816; PID:9511817; EMBL:U11877; NID:9511818; PID:9511819; EMBL:U11878;

J. Biol. Chem. 269, 11065-11072, 1994

A:Title: Structure, genomic organization, and expression of the human interleukin-8 r

A:Reference number: A53611; MUID:94209273; PMID:7512557

A:Accession: A53611

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 6-360 <SPR>

A:Cross-references: GB:M99412; GB:L19593

R:Murphy, P.M.; Tiffany, H.L.

Science 253, 1280-1283, 1991

A:Title: Cloning of complementary DNA encoding a functional human interleukin-8 recep

A:Reference number: A39446; MUID:91368200; PMID:1691716

A:Accession: A39446

A:Status: preliminary; nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 6-360 <MOR>

A:Cross-references: GB:M73969

C:Comment: This receptor, unlike IL8RA, binds several peptides besides interleukin-8,

A:Gene: IL8RA

A:Map position: 2q35-2q35

A:Superfamily: vertebrate rhodopsin

C:Keywords: G protein-coupled receptor; transmembrane protein

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

US-09-830-244b-2 (1-1331) x A53611 (1-360)

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

[illegible]

RESULT 15
SBMOPI
statherin precursor - crab-eating macaque
C:Species: Macaca fascicularis (Crab-eating macaque)
C:Date: 17-Dec-1992 #sequence_revision 21-Jul-1995 #text_change 21-Jan-2000
C:Accession: B32574: A03289
R:Oppenheim, F.G.: Hay, D.I.; Smith, D.J.; Offner, G.D.; Troxler, R.F.
J. Dent. Res. 66, 462-466, 1987
A:Title: Molecular basis of salivary proline-rich protein and peptide synthesis: cellular peptides.

A:Reference number: A92773; PMID:3476566
A:Accession: B32524
A:Molecule type: protein
A:Residues: 159 <DPI>
A:Note: 6-Leu was also found

R:Openheim, F.G.; Offner, G.D.; Troxler, R.F.
 J. Biol. Chem. 257, 9271-9282, 1982
 A:Title: Phosphoproteins in the parotid saliva from the subhuman primate Macaca fascicularis
 of a proline-rich phosphopeptide.
 R:Reference number: A03289; MUID:82265555; PMID:7107568
 A:Accession: A03289
 A:Molecule type: protein
 A:Residues: 20-61 <OP2>
 C:Superfamily: statherin precursor; statherin/histatin signal sequence homology
 C:Keywords: phosphoprotein; saliva
 F:1-25/Domain: statherin/histatin signal sequence homology <SHH>
 F:1-19/Domain: signal sequence #status experimental <SIG>
 F:20-61/Product: statherin #status experimental <MAP>
 F:21,22/Binding site: phosphate (Ser) (covalent) #status experimental

Alignment Scores:

Pred. No.:	1.64	Length:	61
Score:	85.00	Matches:	15
Percent Similarity:	68.18%	Conservative:	0
Best Local Similarity:	68.18%	Mismatches:	7
Query Match:	3.66%	Indels:	0
DB:	1	Gaps:	0

US-09-830-244b-2 (1-1331) x SBWQPI (1-61)

OY	1059	CCTTATCAGCCAGTTCAGAACACACACTATACGACACACACACACACATACCA	1118
Db	40	ProTyrGlnProPheAlaProGlnProLeuTyrProGlnProTyrGlnPro	59
OY	1119	CAATAT	1124
Db	60	GlnTyr	61

Search completed: June 24, 2003, 19:25:33
 Job time : 64 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: June 24, 2003, 19:15:17 ; Search time 21 Seconds

(without alignments)
5257.618 Million cell updates/sec

Title: US-09-830-244b-2

Percent score: 2322
Sequence: 1 ctatgtttttagaatacaag.....gccttttcaatgtcactgt 1331

Scoring table: BIOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL-frame+.n2p.model -DEV-xip
-Q=/cgn2/_USPTO.spool/US09830244/rnmat.24062003.130001.10030/app.query.fasta.1.1479
-DB=SwissProt.40 -OPMT-fastan -SUFFIX=n2p.rsp -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS-bits -SPART=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pcr -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09830244.ecgn.1.1.30.0/rnmat.24062003.130001.10030 -NCPU=6 -ICPU=3
-NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FCGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7
Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	136	5.9	62	1	STAT_HUMAN
2	91.5	4.0	1138	1	P02808 homo sapien
3	90	3.9	853	1	P30337 saccharomyc
4	89	3.8	467	1	DIA3_HUMAN
5	88.5	3.8	353	1	TBB1_PHYPO
6	85.5	3.7	360	1	IL8B_MACMU
7	85.5	3.7	1070	1	IL8B_HUMAN
8	85	3.7	61	1	EXSC_BUCAI
9	85	3.7	320	1	STAT_MACFA
10	85	3.7	567	1	VO2L_FOMPV
11	84	3.6	378	1	NO2M_HAWMI
12	84	3.6	447	1	EDG3_HUMAN
13	83.5	3.6	353	1	TBB1_CYAPA
14	83	3.6	442	1	IL8B_PANTR
15	83	3.6	1037	1	TBB_TRYBR
16	82.5	3.6	362	1	NI20_YEAST
17	82.5	3.6	476	1	ERD1_YEAST
18	81	3.5	835	1	ATXA_BACAN
					Y422_MYCGE
					P47661 mycoplasma

19	80	3.4	445	1	TBB1_COLGR	P22013 colleotric
20	80	3.5	445	1	Y148_MYCPN	P75584 mycoplasma
21	79.5	3.4	482	1	RNF9_HUMAN	O9udv6 homo sapien
22	79.5	3.4	572	1	YB30_YEAST	P38125 saccharomyc
23	79.5	3.4	635	1	ETFL_YABAM	O9q997 yaba monkey
24	79.5	3.4	1329	1	FTSK_ECOLI	P46889 escherichia
25	79.5	3.4	1612	1	RRPO_PPMVY	P89657 pepper mild
26	79.5	3.4	1612	1	RRPO_PPMVY	P29098 pepper mild
27	79	3.4	359	1	IL8B_RAT	P35407 rattus norv
28	79	3.4	447	1	TBB2_PEA	P29501 platan norv
29	78.5	3.4	353	1	IL8B_GORGO	O28422 gorilla gor
30	78	3.4	42	1	STAT_MACFA	P14709 mecate arct
31	77.5	3.4	350	1	IL8A_PANTR	P55920 pan troglod
32	77.5	3.4	483	1	PRH_BUCAI	P57386 buchiera ap
33	77.5	3.4	1237	1	POL4_DROME	P10394 drosophila
34	77	3.3	262	1	MURI_BUCAI	P57619 buchiera ap
35	77	3.3	417	1	IF_HUMAN	P27352 homo sapien
36	77	3.3	443	1	TBB_THAME	O91K18 thalassiosl
37	77	3.3	451	1	TBB1_SOLITU	P46264 solanum tub
38	77	3.3	452	1	TBB2_SOLITU	P46264 solanum tub
39	77	3.3	653	1	GP1L_SCHPO	O14357 schizosacch
40	77	3.3	946	1	YBTF_YEAST	P38250 saccharomyc
41	76.5	3.3	350	1	IL8A_HUMAN	P25024 homo sapien
42	76	3.3	237	1	YPV4_METTE	P29575 methanobact
43	76	3.3	359	1	IL8B_RAT	P35407 rattus norv
44	76	3.3	760	1	ALK1_YEAST	P43633 saccharomyc
45	75.5	3.3	439	1	SYE2_HELPY	O25360 helicobacte

ALIGNMENTS

RESULT 1
STAT_HUMAN
ID STAT_HUMAN STANDARD; PRT; 62 AA.
AC P02808:
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 01-AUG-1991 (Rel. 19, Last annotation update)
DE Statherin precursor.
GN STATH.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP MEDLINE=88074310; PubMed=3502720;
RA Sabatini L., Carlock L., Johnson G., Azen E.;
RT "CDNA cloning and chromosomal localization (4q11-13) of a gene for
RT statherin, a regulator of calcium in saliva.";
RL Am. J. Hum. Genet. 41:1048-1060(1987).
RN [2]
RP MEDLINE=88106506; PubMed=3426601;
RA Dickinson D.P., Ridall A.L., Levine M.J.;
RT "Human submandibular gland statherin and basic histidine-rich peptide
RT are encoded by highly abundant mRNA's derived from a common ancestral
RT sequence.";
RL Biochem. Biophys. Res. Commun. 149:784-790(1987).
RN [3]
RP MEDLINE=90323623; PubMed=2373369;
RA Sabatini L.M., He Y.-Z., Azen E.A.;
RT "Structure and sequence determination of the gene encoding human
RT salivary statherin.";
RL Gene 89:245-251(1990).
RN [4]
RP MEDLINE=77118656; PubMed=838735;
RA Schlesinger D.H., Hay D.I.;
RT "Complete covalent structure of statherin, a tyrosine-rich acidic
RT peptide which inhibits calcium phosphate precipitation from human
RT parotid saliva.";

CC J. Biol. Chem. 252:1689-1695(1977).
RL [5]
RN
RP PARTIAL SEQUENCE OF 1-16.
RX MEDLINE=87309161; PubMed=3476566;
RA Oppenheim F.G., Hay D.I., Smith D.J., Offner G.D., Trowler R.F.;
RT "Molecular basis of salivary proline-rich protein and peptide
RT synthesis: cell-free translations and processing of human and macaque
RT statherin mRNAs and partial amino acid sequence of their signal
RL peptides";
J. Dent. Res. 66:462-466(1987).
CC -I- FUNCTION: SALIVARY PROTEIN THAT STABILIZES SALIVA SUPERSATURATED
CC WITH CALCIUM SALTS BY INHIBITING THE PRECIPITATION OF CALCIUM
CC PHOSPHATE SALTS. IT ALSO MODULATES HYDROXYAPATITE CRYSTAL
CC FORMATION ON THE TOOTH SURFACE.
CC -I- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -I- TISSUE SPECIFICITY: SECRETED BY PAROTID AND SUBMANDIBULAR
CC GLANDS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/>
CC or send an email to license@isb-sib.ch).
CC -----

DR	EMBL:	M18371;	AAA60600.1;	-
DR	EMBL:	M32639;	AAA60593.1;	-
DR	PIR:	A27308;	SBHUP	-
DR	PIR:	A32524;	A32524	-
DR	PIR:	JH0153;	JH0153	-
DR	GeneW:	HGNC:	11369;	STATH.
DR	MIM:	184470;	-	-
KW	Saliva:	Phosphorylation;	Enamel pellicle;	signal.
FT	SIGNAL	1	19	
FT	CHAIN	20	62	
FT	DOMAIN	20	25	
FT	DOMAIN			STATHERIN.
FT	DOMAIN			HYDROXYAPATITE-BINDING (INHIBITS
FT	DOMAIN			CRYSTAL GROWTH).
FT	DOMAIN	38	62	HYDROPHOBIC (INHIBITS PRECIPITATION
FT	MOD_RES	21	21	OF CALCIUM PHOSPHATE SALTS).
FT	MOD_RES	22	22	PHOSPHORYLATION.
SO	SEQUENCE	62 AA;	7304 MW;	CFE32EC235CA3A322 CRC64;

Alignment Scores:	
Pred. No.:	5.14e-06
Score:	136.00
Percent Similarity:	85.71%
Best Local Similarity:	85.71%
Query Match:	5.86%
DB:	1
Gaps:	0
Length:	62
Matches:	24
Conservative:	0
Mismatches:	4
Indels:	0

	US-09-830-24AB-2	(1-1331)	x STAT_HUMAN	(1-62)
OY	1047	TATACCAAGACCCTTATCAGCCAGTTCCAAGAACAACACTATTGCGCACCATACCAA	1106	
Dd	35	 TYIGLYTGELPYOTRGINPROVALPROGLUGINProLeutyrProgInPtyoTrgln	54	
OY	1107	CCACATAACCAACATATACGTTC	1130	
Dd	55	 PROGINTyrglNGinIntyrrThrphe	62	
RESULT 2				
	DPS2_YEAST			
ID	DPS2_YEAST	STANDARD;	PRT:	1138 AA.
AC	P53037;			
DT	01-OCT-1996	(Rel. 34,	Created)	
DR	01-OCT-1996	(Rel. 34,	Last sequence update)	
DD	15-JUL-1998	(Rel. 36,	Last annotation update)	

```
CC - FUNCTION: BINDS TO GTP-BOUND FORM OF RHO AND TO PROFILIN. ACTS IN
CC A RHO-DEPENDENT MANNER TO RECRUIT PROFILIN TO THE MEMBRANE, WHERE
CC IT PROMOTES ACTIN POLYMERIZATION. IT IS REQUIRED FOR CYTOKINESIS,
CC STRESS FIBER FORMATION, AND TRANSCRIPTIONAL ACTIVATION OF THE
CC SERUM RESPONSE FACTOR. DER PROTEINS COUPLE RHO AND SRC TYROSINE
CC KINASE DURING SIGNALING AND THE REGULATION OF ACTIN DYNAMICS (BY
CC SIMILARITY).
CC - ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC - DOMAIN: DERFS ARE REGULATED BY INTRAMOLECULAR GBD-DAD BINDING WHERE
CC RHO-GTP ACTIVATES THE DERFS BY DISRUPTING THE GBD-DAD INTERACTION
CC (BY SIMILARITY).
CC - SIMILARITY: CONTAINS 1 GTPASE-BINDING DOMAIN (GBD).
CC - SIMILARITY: CONTAINS 1 FORNIN HOMOLOG 1 (FH1) DOMAIN.
CC - SIMILARITY: CONTAINS 1 FORNIN HOMOLOG 2 (FH2) DOMAIN.
CC - SIMILARITY: CONTAINS 1 FORNIN HOMOLOG 3 (FH3) DOMAIN.
CC - SIMILARITY: CONTAINS 1 DRF AUTOREGULATORY DOMAIN (DAD).
CC - SIMILARITY: BELONGS TO THE FORNIN HOMOLOG FAMILY. DIAPHANOUS
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL137718; CAB70890.1; ALT_INIT.
DR EMBL; AL354829; CACI17664.1; -.
DR EMBL; AL354829; CACI17665.1; -.
DR Genew; HGNC:15480; DIAPH3.
DR InterPro; IPR003104; FH2.
DR InterPro; IPR000904; Sec7.
DR Pfam; PF02181; FH2; 1.
DR SMART; SM00498; FH2; 1.
DR Coiled coil; Repeat; Alternative splicing.
KW NON_TER 1
FT DOMAIN 304 374 FH1 (PRO-RICH).
FT FT 379 818 FH2.
FT DOMAIN 801 815 DAD.
FT FT 137 167 COILED COIL (POTENTIAL).
FT DOMAIN 241 299 COILED COIL (POTENTIAL).
FT DOMAIN 650 799 COILED COIL (POTENTIAL).
FT DOMAIN 819 822 ARG/TYS-RICH (BASIC).
FT VSPPLIC 656 697 VSQTELEKNIROMGRLOQLKELETFFPPEDIHDKFYTKI
FT P-> GLCFKRHFMALEFAKKRIIPFCIMFPLSHSVF
FT IPINF (IN ISOFORM 2).
FT MISSING (IN ISOFORM 2).
SQ SEQUENCE 853 AA; 98610 MW; B7FA9C745AE1BCD9 CRC64;
Alignment Scores:
Pred. No.: 0.427
Score: 90.00 Length: 853
Percent Similarity: 40.70% Matches: 46
Best Local Similarity: 23.12% Conservative: 35
Query Match: 3.89% Mismatches: 63
Indels: 55
Gaps: 9
US-09-830-244B-2 (1-1331) x DIA3_HUMAN (1-853)
OY 717 GTCTGTGCTTTAAAAACAACAAGAAATGGAGATTCTACTA-----ACT 670
Db 409 ILelyValAlasnglunslustyrYLguAsnValAspldeuducylsylaugluasnTrHr 428
OY 669 GTCTGTGCTCCACAGAGGGGGAATTCAGTAGAGTCGTAGAGAAGAAAACAACCTGCAGGG 610
Db 429 pheCyScysglnglnlyS-----GLuArggarglgunglunspIlegluu 444
OY 609 AAGCACACATCTGGAGATTTCAGATAAGTACCACAAAATAATTATGACRATCTCTTTTGCA 550
Db 445 LysLysSerLeIleLysLysLIleLysLIleLysLIleLysLIleLysLIleLysLIleLysLIleLys 458
```

OY	549	CATGCCCTTTCTCTGTTAAACACTTTCAGTCATTCACAAAGGCTATTGTAAGGTGCAGC	450
Db	458	-----	458
OY	489	AAATCCAGTGTAGATTTTACTGTACAGATTTATTTAACTATGCTGTTACCCATTAAGC	430
Db	459	AspergilluslialaginsluenseSerileheleuenseSerPheargvalProtyrGlu	478
OY	429	GCATTACCAATAAATGGCATT------GTGCACATCAAAAGATCT-----GAA	388
Db	479	GlutlerargmetMetileLeuGluValaLapgluThrArgleuValaginsuermetiIleGln	498
OY	387	AATCCATATGGGACATTTCCCCACAAAGAAAG----AATCCATATGGCTCTTTAAAGG	331
Db	499	Asn-LeuilelyshisLeuProaspGlnGlnLeuAsnSerleuSerGlnPheIysse	518
OY	330	AGATATATAGCAATGGCAAGCAAGATGAATAATTTCTGTCAAAAGAAAGTACTTCTCAAGC	271
Db	518	rgltuYserAsnLeuGlnGlnProGlnGlnPheVal-----Va	531
OY	270	AATCAGGAAAACTGTTCCTCCAAATTTAGGCGCTGGGAGCCACATTACCATATCTCAGAT	211
Db	531	ValmetSerAsnVallysrargLeuArgPro-----ArgleuSerIalI	546
OY	210	ACGATATTATTG----TTTACCTTATGGGGCCCTCAATATATCCCTCATCTGCTA	159
Db	546	eleupheylsLeuGlnPheGlnGlnValaAsnAsnIlelysrProaspIleMet	564

RESULT 4

ID	TTB1_PHYPO	STANDARD;	PRT;	467	AA.
AC	P07436;				
DT	01-APR-1988 (Rel. 07, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DE	Tubulin beta-1 chain.				
DE	BETA AND BETA.				
OS	Physarum polycephalum (Slime mold).				
OC	Eukaryota; Mycetozoa; Myxogastria; Myxogastromycetidae; Physarida;				
NCBI	Physarum.				
NCBI	TaxID=5791;				
RN	(1)				
RC	SEQUENCE FROM N.A.				
RC	STRAIN=LU352;				
RA	Paul E.C.A., Buchsacher G.L. Jr., Cunningham D.B., Dove W.F.,				
RA	Burland T.G.;				
RL	Submitted (Mar-2000) to the EMBL/GenBank/DBJ databases.				
RP	(2)				
RP	SEQUENCE OF 1-445 FROM N.A.				
RC	STRAIN=LU352;				
RX	MEDLINE=92211323; PubMed=1556511;				
RA	Paul E.C.A., Buchsacher G.L. Jr., Cunningham D.B., Dove W.F.,				
RA	Burland T.G.;				
RT	"Preferential expression of one beta-tubulin gene during flagellate				
RT	development in Physarum.";				
RL	J. Gen. Microbiol. 138:229-238(1992).				
RN	(3)				
RP	SEQUENCE OF 4-207 FROM N.A.				
RC	STRAIN=CL;				
RX	MEDLINE=88271316; PubMed=3391166;				
RA	Werenskiold A.K., Poetsch B., Haugli F.;				
RT	"Cloning and expression of a beta tubulin gene of Physarum				
RT	polycephalum.";				
RL	Eur. J. Biochem. 174:491-495(1988).				
RN	(4)				
RP	PARTIAL SEQUENCE OF 1-217; 234-262 AND 277-286.				
RP	MEDLINE=87080317; PubMed=3539596;				
RA	Slingshofer-Morra M., Clayton L., Dawson P., Gull K., Little M.;				
RT	"Amino-acid sequence data of beta-tubulin from Physarum polycephalum				
RT	myamebae.";				
RL	Eur. J. Biochem. 161:669-679(1986).				
CC	-1- FUNCTION: TUBULIN IS THE MAJOR CONSTITUENT OF MICROTUBULES. IT				
CC	BINDS TWO MOLES OF GTP, ONE AT AN EXCHANGEABLE SITE ON THE BETA				

```

CC      CHAN AND ONE AT A NONEXCHANGEABLE SITE ON THE ALPHA-CHAIN.
CC      -1- SUBUNIT - DIMER OF ALPHA AND BETA CHAINS.
CC      -1- SUBCELLULAR LOCATION: MITOSIS IN THE SLIME MOLD PLASMODIUM DIFFERS
CC      FROM THE PROCESS IN MANY EUKARYOTES. THE TUBULIN CHAINS MUST BE
CC      TRANSPORTED TO THE NUCLEI FOR INTRANUCLEAR ASSEMBLY OF THE
CC      SPINDLE.
CC      -1- DEVELOPMENTAL STAGE: BETA IS PREFERENTIALLY EXPRESSED IN
CC      FLAGELLATE AND BETB IN AMOEBA.
CC      -1- SIMILARITY: BELONGS TO THE TUBULIN FAMILY.
CC      -----
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL collaboration
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).

```

DR	EMBL: X12371; CA930932.1; -;	
DR	PIR: A44848; A44848.	
DR	PIR: A25342; A25342.	
DR	PIR: S02532; S02532.	
DR	InterPro: IPR002453; Beta_tubulin.	
DR	InterPro: IPR000217; Tubulin.	
DR	InterPro: IPR003008; Tubulin_FtsZ.	
DR	Pfam: PF00091; tubulin_1.	
DR	PRINTS: PR01161; TUBULIN_1.	
DR	PROSITE: PS00227; TUBULIN_1.	
DR	PROSITE: PS00228; TUBULIN_BAUTOREG.1.	
DR	Microtubules: GTP-binding; Multigene family; Nuclear protein.	
KW	Microtubules; GTP-binding; Multigene family; Nuclear protein.	
FT	NE_BIND 140 146	
FT	VARIANT 39	
FT	VARIANT 283	
FT	CONFLICT 165 165	
FT	CONFLICT 196 196	
FT	CONFLICT 238 238	
SQ	SEQUENCE 467 AA; 52134 MW; BA2C330A65FDC964 CRC64;	

Alignment Scores:	
Pred. No.:	0.505
Score:	89.00
Percent Similarity:	39.81%
Best Local Similarity:	21.36%
Query Match:	3.83%
DB:	1
US-09-830-244B-2 (1-1331) x TBB1_PHRPO (1-467)	
Length:	467
Matches:	44
Conservative:	44
Mismatches:	76
Indels:	48
Gaps:	8

Y7 506 TTTTCAAGATCTTTTGATGTGGCACTAATGCATTATGTGAATCCGATTATGGGAATAC 44
 Db 159 TTTTCAAGATCTTTTGATGTGGCACTAATGCATTATGTGAATCCGATTATGGGAATAC 44
 Y 446 AGCATAGTAAATAAATCTTACAGAAATCTACAC----- 481
 Db 179 ValValGluProGlyrAsnAlaThrIleuSerValIleHisGluLeuValGluAsnAlaAspGlu 199
 Y 482 -----TTGGATTTCCTGCACCTCTACCAATAGCCTTTTGATGACCTGAAGATGTTA 532
 Db 199 ValMetCysIleLeuPaspGlnGluAlaLeuValrAspIleSerPheArgThrIleuLeuThr 218
 Y 533 ACAGAGAAAGAGGATGCTGCAGAAACAGAAAGTAAATATTTTTCGTACTTATTCGA 592
 Db 219 ThrProThrTyrGlyAspLeuAsnHisIleuValSerAlaVal-----MetSer 234
 Y 593 AATCCAAAGATGCTGC-----TTCCCTGCAGGTTGTTTTCCTTTCACATCC----- 640
 Db 235 GlyIleThrCysCysLeuArgPheProGlyIleuLeuAsnSerAspLeuArgIleuLeuAla 254
 Y 641 ---TCATTGAATCCCTCTGGGAGACAGACAGACAGTGTAGAACTCCATTCCTTTCTTTGTT 697
 Db 255 ValAspLeuIleProPhe-----ProArgIleuHisPheIleLeu 267

```

OY      698  TTGTTT-----TTAAAGACAGACACTGTGTCTCAAAAAAAGCATTATTCATTA-----748
Db      268  ValIeIhpehlaProIeuthrSerargIySerValGIyTyarGserLeuthValPro 287
OY      749  -----TMACTATTATTAGAGCCCTAAT-----772
Db      288  GluLeuthrGlnImethrheaspIalysImethrIaIaIaSeraspProargHis 307
OY      773  -----TCTTACTCAAGCAGCACTGTTTTTTTTTAAACAGTTAAAGTACGATGCTAAC 826
Db      308  GlYarGtyrLeuthrIaSerIaIaSerIaIaMetrheargIyargMetserThrIySGluValaSP 327
OY      827  AGACAAATATTTCTGTGATCAGATAGTCCCTCGTCAAGGTACCAATATGGTTTCATTAAG 886
Db      328  GluGlnMetLeuAsnValGln-----AsnIysAsnSerSerrTyrPheValGlu 343
OY      887  TGGGAAGAAAGAACAGCAT 904
Db      344  TrpIleProAsnAsnIle 349

```

RESULT 5	IL8B_MACMU	STANDARD:	PRT:	353 AA.
ID	IL8B_MACMU			
AC	028519;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	High affinity interleukin-8 receptor B (IL-8R B) (CXCR-2) (Fragment).			
GN	IL8RB OR CXCR2.			
OS	Macaca mulatta (Rhesus macaque).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;			
OX	Cercopithecoidea; Macaca.			
OX	NCBI_TaxID:9544;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	MEDLINE-9617511; PubMed-9110929;			
RA	Alvarez V., Coto E., Setten F., Gonzalez S., Gonzalez-Roces S.,			
RA	Lopez-Larrea C.;			
RL	"Characterization of interleukin-8 receptors in non-human primates.";			
RL	Immunogenetics 43:261-267(1996).			
CC	-1- FUNCTION: RECEPTOR TO INTERLEUKIN-8, WHICH IS A POWERFUL			
CC	NEUTROPHILS CHEMOTACTIC FACTOR. BINDING OF IL-8 TO THE RECEPTOR			
CC	CAUSES ACTIVATION OF NEUTROPHILS. THIS RESPONSE IS MEDIATED VIA A			
CC	G-PROTEIN THAT ACTIVATES A PHOSPHATIDYLINOSITOL-CALCIUM SECOND			
CC	MESSENGER SYSTEM. THIS RECEPTOR BINDS TO IL-8 WITH A HIGH AFFINITY			
CC	AND TO GRO/MGSA AND NAP-2 ALSO WITH A HIGH AFFINITY.			
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein.			
CC	-1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	at the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/).			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; X91116; CAA62565.1;			
DR	HSSP; P34996; 1DDO.			
DR	InterPro; IPR000276; GPCR_Rhodopsn.			
DR	InterPro; IPR000832; GPCR_secretln.			
DR	Pfam; PF00001; 7tm1_1; 1.			
DR	PROSITE; PS00237; G-PROTEIN_RECPE_F1_1; 1.			
DR	PROSITE; PS50262; G-PROTEIN_RECPE_F1_2; 1.			
KW	G-protein coupled receptor; Transmembrane; Glycoprotein;			
KW	Chemotaxis.			
FT	NON_TER	1	45	
FT	DOMAIN	<1	45	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	46	72	1 (POTENTIAL).
FT	DOMAIN	73	81	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	82	102	2 (POTENTIAL).
FT	DOMAIN	103	117	EXTRACELLULAR (POTENTIAL).

FT	TRANSMEM	118	139	3	(POTENTIAL).
FT	DOMAIN	140	160	3	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	161	180	4	(POTENTIAL).
FT	DOMAIN	181	205	5	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	206	228	5	(POTENTIAL).
FT	DOMAIN	229	248	6	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	249	270	6	(POTENTIAL).
FT	DOMAIN	271	291	7	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	292	312	7	(POTENTIAL).
FT	DOMAIN	313	>353	8	CYTOPLASMIC (POTENTIAL).
FT	DISULFID	116	193		BY SIMILARITY.
FT	CARBOHYD	19	19		N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	NON_TER	353.	353		
SO	SEQUENCE	353	AA: 39947	MM: EC8B38130657C713	CRG64;
Alignment Scores:					
Pred. No.:		0.551			
Score:		88.50		Length:	353
Percent Similarity:		37.84%		Matches:	59
Best Local Similarity:		22.78%		Conservative:	39
Query Match:		3.83%		Mismatches:	87
DB:		1		Indels:	75
				Gaps:	10

QY 1187 AGTCGATATTTGCCAATCAAGCCATCATATATCATGTCCAGCTACTGATC-----ATA 1134
 Db 73 SerArgValGlyArgSerValThrSpValIleuLeuAsnLeuAlaLeuAlaAspLeu 92
 QY 1133 TTAAACGATATTTGTTGCTATTTGGCTTGGTGTATGGTTGCTGCAGTAACTGGTGTGGG 1074
 Db 93 LeuPheAlaLeuThrLeuProIleThrIleAlaSerLysValAsnGlyTrpIlePheGly 112
 QY 1073 AACCTGCTGATTAAGCCTCTTGTGTAAAGCAATATCCACAATTAAGATATTAACTTT 1014
 Db 113 ThrPheLeu-----CysLysValIleSerLeuLeuLysGluValAsnPhe 127
 QY 1013 GGAAGCTAAGGGATTTAATATACACAGTAAGAAGGAGGTTTTTTTTCTTAAGGAG 954
 Db 128 -----TyrSerGly----- 130
 QY 953 ACCAGAGTTTATTATTACTCAATCAAGCTCCCA-AAAAGTTACTTTAAATGCTGTTT 895
 Db 131 -----IleLeuLeuAlaCysIleSerValAspArgTyrLeuAlaIleHis 147
 QY 894 TCTTCCCACTTATGAACAACATTTCTACTGTTCACAGGGAGCATCTGCATCACAAT 835
 Db 148 AlaThrArg-----ThrLeuThrGlnLysArgTyrLeuValLysPhe 161
 QY 834 ATTTGTT-----CTGTTCACATCACTACTTAAGCTTTTAAAAAAA 796
 Db 162 IleCysLeuSerIleTrpGlyLeuSerLeuLeuAlaLeuProValLeuLeuPheArg 181
 QY 795 AAAACAGTGCCTTCAGATTAGAATAATAGGCGCTAATAGATGTATTAATGATAATGT 736
 Db 182 ArgThrValIlyrSerSerAsnValSerProIa-----CysTyrGlnLysP- 196
 QY 735 CCTTTTTTTGGACAGAGCTCTGTCTTTAAAAAACAAACAAGAATGAGAGTTCTA 676
 Db 197 -----MetGlyAsnAsnThrAlaAsnTrpArgMetLeu 207
 QY 675 CTACTGTCCTGTGCTCCACAGAGGGATTCATGAGAGAT----- 637
 Db 208 LeuArgIleLeuProGlnSerPheGlyPheIleValIProLeuLeuIleMetLeuPheCys 227
 QY 636 -----CGTAAAGAGAAAAACAACCTCCAGGGGAAGACAGACACTTGGATT 592
 Db 228 TyrGlyPheThrLeuArgThrLeuPheLysAlaHisMetGlyGlnLysHisArgAlaMet 247
 QY 591 CAGATTAAGTACCAAAAAATATTACTTACTCTTTCTGCACAACTGGCTCTTCTGTT 532
 Db 248 ArgValIlePheAlaValValLeuIleLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuVal 267

```

0Y 531 AACACTTCA-----GTCATTCAAAAGCGTATTGGTAGA 498
Db 267 ILdeuLeuAlaAspThrLeuMetCArgThGlnValIleGlnGluThrCysGlnUarg 285
: |||::: |||||::: |||
RESULT 6
IL8B_HUMAN STANDARD: PRT: 360 AA.
AC ID IL8B_HUMAN P25025;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE High affinity Interleukin-8 receptor B (IL-8R B) (CXCR-2) (GRO/MGSA
GN receptor) (IL-8 receptor type 2) (CDw128b).
DE IL8RB OR CXCR2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Plimates; Carnivora; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN 1)
RP SEQUENCE FROM N.A.
RX MEDLINE=91368200; PubMed=1891716;
RA Murphy P.M., Tiffany H.L.;
RT Cloning of complementary DNA encoding a functional human
RT Interleukin-8 receptor.
RL Science 255:1280-1283(1991).
RN 12)
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=93205012; PubMed=8384312;
RA Cerretti D.P., Kozlosky C.J., Vanden Bos T., Nelson N., Gearing D.P.,
RA Beckmann M.P.;
RT "Molecular characterization of receptors for human Interleukin-8,
RT GRO/melanoma growth-stimulatory activity and neutrophil activating
RT peptide-2."
RL Mol. Immunol. 30:359-367(1993).
RN 13)
RP SEQUENCE FROM N.A.
RX MEDLINE=94209273; PubMed=7512557;
RA Sprenger H., Lloyd A.R., Lautens L.L., Bonner T.I., Kelvin D.J.;
RT Structure, genomic organization, and expression of the human
RT Interleukin-8 receptor B gene."
RL J. Biol. Chem. 269:11065-11072(1994).
RN 14)
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=95014476; PubMed=7929358;
RA Ahuja S.K., Shetty A., Tiffany H.L., Murphy P.M.;
RT "Comparison of the genomic organization and promoter function for
RT human Interleukin-8 receptors A and B."
RL J. Biol. Chem. 269:26381-26389(1994).
RN 15)
RP CHARACTERIZATION.
RX MEDLINE=92355587; PubMed=1379593;
RA Lee J., Horuk R., Rice G.C., Bennett G.L., Camerato T., Wood W.I.;
RT "Characterization of two high affinity human Interleukin-8
RT receptors."
RL J. Biol. Chem. 267:16283-16287(1992).
RN 16)
RP FUNCTION: RECEPTOR TO INTERLEUKIN-8, WHICH IS A POWERFUL
RN NEUTROPHILS CHEMOTACTIC FACTOR. BINDING OF IL-8 TO THE RECEPTOR
RN CAUSES ACTIVATION OF NEUTROPHILS. THIS RESPONSE IS MEDIATED VIA A
RN G-PROTEIN THAT ACTIVATE A PHOSPHATIDYLINOSITOL-CALCIUM SECOND
RN MESSENGER SYSTEM. THIS RECEPTOR BINDS TO IL-8 WITH A HIGH AFFINITY.
RN AND TO GRO/MGSA AND NAP-2 ALSO WITH A HIGH AFFINITY.
RN 17) SUBCELLULAR LOCATION: Integral membrane protein.
RN 18) SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
RN 19) DATAABAS: NAME=PROM; NOTE=CD guide CDw128b entry;
RN WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cdw128b.htm".
RN 20)
RN This SWISS-PROT entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL outstation -
RN the European Bioinformatics Institute. There are no restrictions on its
RN use by non-profit institutions as long, as its content is in no way
RN modified and this statement is not removed. Usage by and for commercial
RN entities requires a license agreement (See http://www.isb-sib.ch/announce/)

```

CC or send an email to license@lsb-sib.ch).
 CC -----

```

DR EMBL; M73969; AAA83148.1; -
DR EMBL; M94582; AAA36108.1; -
DR EMBL; M99412; AAC14460.1; -
DR EMBL; L19593; AAB59437.1; -
DR EMBL; U11869; AAB60656.1; -
DR PIR; A39446; A39446; -
DR PIR; A53611; A53611; -
HSP; P34996; 1DD; -
GeneW; HENC; 6027; IL8RB.
DR MIM; 146928; -
DR InterPro; IPR000276; GPCR_Rhodopsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRRHOOPS.
DR PROSITE; PS00237; G-PROTEIN_RECP_FL_1; 1.
DR PROSITE; PS0262; G-PROTEIN_RECP_FL_2; 1.
KW G-protein coupled receptor; Transmembrane; glycoprotein;
KM Chemotaxis.
FT DOMAIN 1 48 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 49 75 1 (POTENTIAL).
FT DOMAIN 76 84 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 85 105 2 (POTENTIAL).
FT DOMAIN 106 120 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 121 142 3 (POTENTIAL).
FT DOMAIN 143 163 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 164 183 4 (POTENTIAL).
FT DOMAIN 184 208 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 209 231 5 (POTENTIAL).
FT DOMAIN 232 251 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 252 273 6 (POTENTIAL).
FT DOMAIN 274 294 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 295 315 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 316 360 7 (POTENTIAL).
FT CARBOHYD 316 360 CYTOPLASMIC (POTENTIAL).
FT DISULFID 22 22 N-LINKED GLCNAC. ) (POTENTIAL).
SQ SEQUENCE 360 AA; 40759 MW; 564F0A4B8BC0A197 CRC64;

Alignment Scores:
Pred. No.: 1 13 Length: 360
Score: 85.50 Matches: 58
Percent Similarity: 37.84% Conservative: 40
Best Local Similarity: 22.39% Mismatches: 87
Query Match: 3.70% Indels: 75
DB: 1 Gaps: 10

US-09-830-244B-2 (1-1331) x IL8B_HUMAN (1-360)
QY 1187 AGTCGATATTGCCANTCAAGCCCATATATCATCTCGTACTGATG-----ATA 1134
Db |||||::: ::||| : : : : : ||| : : : : :
QY 76 SerArgValGlyArgSerValThrAspAlaIleuLeuAsnLeuAlaLeuAlaAspLeu 95
Db |||||::: ::||| : : : : : ||| : : : : :
QY 1133 TTAAGCATTATTTGGTGATGTGGTGTTGGTATAGTTGCCGATATAGTGGTGTCTGG 1074
Db |||||::: ::||| : : : : : ||| : : : : :
QY 96 LeuhenAlaLeuThrLeuProIleThrAlaIleSerLysValAsnGlyTrpIleebecUy 115
Db |||||::: ::||| : : : : : ||| : : : : :
QY 1073 AACTGGCTGATAAAGCTCTTTGGTATAAAGCAATATCCAGAATAATAGATATTAACCTT 1014
Db |||||::: ::||| : : : : : ||| : : : : :
QY 116 ThrheLeu-----CylbLysValAlaSerLeuLeuLysGLuValasnphe 130
Db |||||::: ::||| : : : : : ||| : : : : :
QY 1013 GAACGTAAGGGGATATAAATGACACACATMAAGTGGGAAGGTTTTTTTCTTAAGCGAG 954
Db |||||::: ::||| : : : : : ||| : : : : :
QY 131 -----TyserGly----- 133
Db |||||::: ::||| : : : : : ||| : : : : :
QY 953 ACCAGAGTTTATATTATTACTCAATCAGCTCCCA-AAAGTACTTTAAATGCCTGTT 895
Db |||||::: ::||| : : : : : ||| : : : : :
QY 134 -----IleuLeuLeuAlaIcysIleSerValaspArgIryLeuAlaIleValHis 150
Db |||||::: ::||| : : : : : ||| : : : : :
QY 894 TCCTCCACACTTMTGAACACACTTTCCTACTGTTCGACAGGGGACTATCTGATCAGAAAT 835
Db |||||::: ::||| : : : : : ||| : : : : :
QY 151 AlathrArg-----ThrLeuThrInLysArgIryLeuValLysphe 164
Db |||||::: ::||| : : : : : ||| : : : : :
```

OY 834 ATTGCT-----CTGTGACATGACTTAACTGTTAAAAA 796
 Db 165 IlecylseuSerIleTTPGlyLeuSerLeuLeuLeuLeuProValLeuLeuPheArg 184
 OY 795 AAACAGTGCCTTCAGATTAAGAAATTAGGGCTTAATAGATGTTATGATTAATGT 736
 Db 185 ArghValIlySerSerAsnValSerProAla-----CysTYUGLAsp----- 199
 OY 735 CCTTTTTCGAGACAGAGTCTCTGTAAAAAACAAGAAAGAAATGAGAGTTCTA 676
 Db 200 -----MetGlysnAsnThrAlaAsnTPArgMetLeu 210
 OY 675 TTAAGTCTCTGCTGCTCCAGAGGGATTAATGAGAT----- 637
 Db 211 LeuArgIleLeuProGlnSerPheGlyPheIleValProLeuLeuIleMetLeuPheCys 230
 OY 636 -----CSTAAGAGAAAGAAACACCTGAGGGAGAGCAGCATTTGATTT 592
 Db 231 TyrGlyPheThrLeuArgThrLeuPheIlysalHisMetGlyGlnYshIsArgAlaMet 250
 OY 591 CAGATTAAGTACCAAAAAATATTAGCTATCTTTCTGACAGACATGCTCTCTGTT 532
 Db 251 ArgValIlePheValIleValIlePheLeuLeuCysTrp-LeuProTyrAsnLeuVal 270
 OY 531 AACACTTCA-----GTCATTCAAAAGCCTATTGCTAGA 498
 Db 270 IleuLeuAlaAspThrLeuMetArgThrGlnValIleGlnGluThrCysGlnArg 288
 RESULT 7
 EX5C_BUCAI STANDARD: PRT; 1070 AA.
 AC P57528:
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Exoexyribonuclease V gamma chain (Ec 3.1.11.5).
 GN RECC OR BU453.
 OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
 symbiotic bacterium).
 OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
 OX NCBI_TaxID=118099;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Tokyo 1998;
 RX MEDLINE=20445173; PubMed=10993077;
 RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
 RT "Genome sequence of the endocellular bacterial symbiont of aphids
 Buchnera sp. APS.";
 RL Nature 407:81-86(2000).
 CC -1- FUNCTION: EXHIBITS A WIDE VARIETY OF CATALYTIC ACTIVITIES
 INCLUDING ATP-DEPENDENT EXONUCLEASE, ATP-STIMULATED ENDONUCLEASE,
 CC ATP-DEPENDENT HELICASE AND DNA-DEPENDENT ATPASE ACTIVITIES
 CC (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: Exonucleolytic cleavage (in the presence of
 CC ATP) in either 5'- to 3'- or 3'- to 5'-direction to yield 5'-
 CC phosphooligonucleotides.
 CC -1- SUBUNIT: CONSIST OF THREE SUBUNITS; RECB, RECC AND RECD
 CC (BY SIMILARITY).
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AP001119; BAB1351.1;
 KW Hydroxylase; Nuclease; Exonuclease; Endonuclease; Helicase; DNA repair;
 SO Complete proteome.
 SEQUENCE 1070 AA; 128578 MW; 8F644C8487981AF CRC64;
 Alignment Scores:

Pred. No.: 1.29 Length: 1070
 Score: 85.50 Matches: 78
 Percent Similarity: 38.52% Conservative: 63
 Best Local Similarity: 21.31% Mismatches: 102
 Query Match: 3.68% Indels: 123
 DB: 1 Gaps: 22
 US-09-830-244B-2 (1-1331) x EX5C_BUCAI (1-1070)
 OY 261 TTCCGATGTCGTTGAGAGTACTTCTTTGACAGAAAT---TTCATTCGCTTGCC 317
 Db 105 PheGlnAsnGlySerIlyLysLysAspMetIleLysLysPheLysPheSerPheLeu 124
 OY 318 ATTGCTATATTCCTCC-----CTTATAGAGCATGATTTCTTCTTTGTCG 368
 Db 125 AlaSerIlePheLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 142
 OY 369 GAATGTCCTCCATTGACATTTTCAGATCTTTGATGTCACATAATGCCATTATGTAATG 428
 Db 143 GlnIleGluLys---AsnIleSerIle-PheAspLysAsn-----GlnGlnTrpGln 159
 OY 429 CGGTATGTCGATATACAGCATAGTTAAATTAAGCTGTTACAGTAATCTACACTTGATTT 488
 Db 159 eLysLeuTrp---MetGlu-IleIleHisAsnThrLysLysIleAsnGlnSerAsp- 177
 OY 489 TGCTGCACCTTCACCAATAGCTTTTGATGACTGAAAGTGTAAACAGAAAGAGCAT 548
 Db 178 ----HisPheAlaAsnLeuPheTyrAsnIleGlnIlyLysLysLysLysLysLysLys 194
 OY 549 GTCTGACAGAAAGATAGCTAATTAATTTTGGTACTTATCTGAATCCAGATGCTGCT 608
 Db 195 --IleLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 200
 OY 609 TCCCTGCAGGTTGTTTCTTCTTACATCTCATTTGAATCC-----CTCT 656
 Db 200 euprolYsArgPhePheIleIleSerPheSerMetAsnProSerTyrIleLysIleP 220
 OY 657 GGGAGCACAGACAGATTAGTAGACTCTCCATTC---TTTGTTTCTTTTAAAGACAG 713
 Db 220 heGlnAsnIleSerIleTyrThrAspIleTyrPheLeuTyrIleThrProPheLysLys 240
 OY 714 AGACTCTGTCCAAAAAAAGACATTATTCATTATACATCTTATAGACCCCTAATTT 773
 Db 240 snIlePhe-----Asn 244
 OY 774 TCTTATCTGAAGCAGCTGTTTCTTTTAAACAGTTAAGTACTGATCTACAGACAAA 833
 Db 244 heIleGlnAspAsnLysIlePhe-----ThrAsp----- 253
 OY 834 TATTTCGATCAGATAGTATGCTCCCTGTCACAGTAGCAATGTTGTTCAATAAGGGAG 893
 Db 254 -----IleLysIleGlu 258
 OY 894 AAAACAGATTTTAAAGTACTTTTGGAGACTGATTTAGTAATATAAACTGCT 953
 Db 258 yLysAsnIleLeuAsn-----AspSerLeuIleThrLeuTrp 271
 OY 954 CTCCTTAAGAAAAAACCCTTCCACCTTTACTGTCGATTTATATCCCTTAGTTC 1013
 Db 271 LysGlnTrp-GluLys-----IleTyrThrLeuLysLysSer 284
 OY 1014 AAAGTTAATATTTATTTCTGATATGCTTT-----TATACCAAGAGCCCTAT 1064
 Db 285 Lys 302
 OY 1065 CAGCCAGTTCGAGAACACACATATGACACACATACCAACAC----- 1110
 Db 302 uSerGlnIleLysAsnAsnPhePheAsnAspSerGlnPheThrArgLysLysArgPheLe 322
 OY 1111 -----AATACCAATATATGTTTATATATCATCACTAGTACAGACATGATTAAT 1163
 Db 322 uGlnIleSerAspHisSerIleSerIleAsnIleCysPheAsnLysLysAsn----- 339

OY 1164 GAGGCTTGATGGCAATACGACTTCTACATCATATTCATCTTTCATACCAATAC 1223
 Db 340 -----GluilegualleuTyT-----G1 346
 OY 1224 ACTACTACACTTTTGTGATCATCTAAGACGATGCAATGAAACCTTAATTT 1233
 Db 346 ulyleueneulerheleu-----Asnclunserseilerlyspgilyaspil 363
 OY 1284 ACTGATACT 1293
 Db 363 evalvalthr 366

RESULT 8

STAT_MACFA STANDARD: PRT: 61 AA.
 AC P02809; 21-JUL-1986 (Rel. 01, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DE 15-DEC-1998 (Rel. 37, Last annotation update)
 GN Statherin precursor.

OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecinae; Macaca.
 NCBI_TaxID=9541;
 RX PARTIAL SEQUENCE OF 1-19.
 MEDLINE=87309161; PubMed=3476566;
 RA Oppenheim F.G., Hay D.I., Smith D.J., Offner G.D., Troxler R.F.;

RT Molecular basis of salivary proline-rich protein and peptide
 RT synthesis: cell-free translations and processing of human and macaque
 RT statherin mRNAs and partial amino acid sequence of their signal
 RT peptides.";
 RL J. Dent. Res. 66:462-466(1987).
 RN [2]
 RP SEQUENCE OF 20-61.
 RA MEDLINE=82265555; PubMed=7107568;
 RX Oppenheim F.G., Offner G.D., Troxler R.F.;

RT Phosphoproteins in the parotid saliva from the subhuman primate
 RT Macaca fascicularis. Isolation and characterization of a proline-rich
 RT phosphoglycoprotein and the complete covalent structure of a proline-
 RT rich phosphopeptide.";
 RL J. Biol. Chem. 257:9271-9282(1982).
 RN [1]
 RP FUNCTION: SALIVARY PROTEIN THAT STABILIZES SALIVA SUPERSATURATED
 CC WITH CALCIUM SALTS BY INHIBITING THE PRECIPITATION OF CALCIUM
 CC PHOSPHATE SALTS. IT ALSO MODULATES HYDROXYAPATITE CRYSTAL
 CC FORMATION ON THE TOOTH SURFACE.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- TISSUE SPECIFICITY: SECRETED BY PAROTID AND SUBMANDIBULAR
 CC GLANDS.

DR PIR: A03289; SBMOP1.
 KM Saliva; Phosphorylation; Enamel pellicle; Signal.
 FT SIGNAL 1 19
 FT CHAIN 20 61
 FT MOD_RES 21 21
 FT MOD_RES 22 22
 FT DOMAIN 20 25
 FT DOMAIN 37 61

FT STATHERIN.
 FT PHOSPHORYLATION.
 FT PHOSPHORYLATION.
 FT HYDROXYAPATITE-BINDING (INHIBITS
 FT CRYSTAL GROWTH).
 FT HYDROPHOBIC (INHIBITS PRECIPITATION
 FT OF CALCIUM PHOSPHATE SALTS).
 FT CARBOHYD 175
 SQ SEQUENCE 61 AA; 7452 MW; 64241AAB5641ASB CRC64;

Alignment Scores:

Pred. No.: 1.03 Length: 61
 Score: 85.00 Matches: 15
 Percent Similarity: 68.18% Conservative: 0
 Best Local Similarity: 68.18% Mismatches: 7
 Query Match: 3.66% Indels: 0
 DB: 1 Gaps: 0

US-09-830-244B-2 (1-1331) x STAT_MACFA (1-61)

OY 1059 CCTATACGCGAGTTCAGAACACACTATACGACACACCAACCAACCAATACCA 1118
 Db 40 ProtyrlnProhealaproInProleutyProInProtyrlnProtyrlnPro 59
 OY 1119 CAATAT 1124
 Db 60 GlnTyT 61

RESULT 9

V021_F0WMPV STANDARD: PRT: 320 AA.
 ID V021_F0WMPV
 AC Q93510;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE 16-OCT-2001 (Rel. 40, Last annotation update)
 GN G-protein coupled receptor homology FPV021.
 GN FPV021.
 OS Fowlpox virus (FPV).
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
 OC Avipoxvirus.
 NCBI_TaxID=10261;
 RX MEDLINE=20193820; PubMed=10729156;
 RA Alonso C.L., Tulman E.R., Lu Z., Zsak L., Kutish G.F., Rock D.L.;

RT "The genome of fowlpox virus."
 RL J. Virol. 74:3815-3831(2000).
 RN [1]
 RP SEQUENCE FROM N.A.
 MEDLINE=20193820; PubMed=10729156;
 RA Alonso C.L., Tulman E.R., Lu Z., Zsak L., Kutish G.F., Rock D.L.;

CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

DR EMBL: AF198100; AAF44365.1;
 DR InterPro: IPR000276; GPCR_Rhodopsn.
 DR Pfam: PF00001; 7tm1.1.
 DR PRINTS: PR00237; GPCR_RHODOPSN.
 DR PROSITE: PS00337; G-PROTEIN-RECEP_F1_1; FALSE_NEG.
 DR PROSITE: PS50262; G-PROTEIN-RECEP_F1_2; 1.

KW G-protein coupled receptor; transmembrane; Glycoprotein.
 FT DOMAIN 1 18
 FT TRANSMEM 19 39
 FT DOMAIN 40 52
 FT TRANSMEM 53 73
 FT DOMAIN 74 91
 FT TRANSMEM 92 112
 FT DOMAIN 113 133
 FT TRANSMEM 134 154
 FT DOMAIN 155 188
 FT TRANSMEM 189 209
 FT DOMAIN 210 222
 FT TRANSMEM 223 243
 FT DOMAIN 244 260
 FT TRANSMEM 261 281
 FT DOMAIN 282 320
 FT DISULFID 89 167
 FT CARBOHYD 175
 SQ SEQUENCE 320 AA; 37808 MW; 5F082B9A1C6161958 CRC64;

Alignment Scores:

Pred. No.: 1.26 Length: 320
 Score: 85.00 Matches: 56
 Percent Similarity: 39.51% Conservative: 40
 Best Local Similarity: 23.05% Mismatches: 59
 Query Match: 3.66% Indels: 88
 DB: 1 Gaps: 14

US-09-830-244B-2 (1-1331) x V021_F0WMPV (1-320)

[illegible]

[illegible]

```

QY      1037 ATATTGGTTTATTAACAAAGGCTTATCATGCCAGGTCCAGAACAACACTATACGCACA   1096
Db      217 rglIeTYrPhelValLysSerSeraArglySvalAlasn-----HISA    233
QY      1097 ACCATACCAACACATATACCACAAATVACGGTTTATATCATC   1140
Db      233 snasnserrguaryselmeretalaleuLeuarThyValValle   247

RESULT 12
TBBL_CYPAP
ID      TBBL_CYPAP          STANDARD;             PRT;       447 AA.
AC      Q9ZSW1;
DT      16-OCT-2001 (Rel. 40, Created)
DE      16-OCT-2001 (Rel. 40, Last sequence update)
DE      16-OCT-2001 (Rel. 40, Last annotation update)
GN      Tubulin beta-1 chain (Beta-1 tubulin).
OS      Cyanophora paradoxa.
OC      Eukaryota; Glaucocystophyceae; Cyanophoraceae; Cyanophora.
OX      NCBI_TaxId=2762;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=UB555 / Pringsheim;
RA      Keeling P., Hink-Schauer C., Loeffelhardt W.;
RT      "Phylogenetic analysis of beta.tubulins.";
RL      Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
CC      - FUNCTION: TUBULIN IS THE MAJOR CONSTITUENT OF MICROTUBULES. IT
CC      BINDS TWO MOLES OF GTP, ONE AT AN EXCHANGEABLE SITE ON THE BETA
CC      CHAIN AND ONE AT A NONEXCHANGEABLE SITE ON THE ALPHA-CHAIN.
CC      - SIMILARITY: BELONGS TO THE TUBULIN FAMILY.
CC
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outpost at
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC
DR      EMBL: AF092952; AACD03712.1; -
DR      InterPro: IPR002453; Beta_tubulin.
DR      InterPro: IPR000217; Tubulin.
DR      InterPro: IPR003008; Tubulin_FtsZ.
DR      Pfam: PF00091; tubulin_1.
DR      PRINTS: PRO1161; TUBULIN.
DR      PROSITE: PS00227; TUBULIN.
DR      PROSITE: PS00228; TUBULIN_B_AUTOREG.
KW      Microtubules; GTP-binding; Multigene family.
NP_BIND 140 146 GTP (POTENTIAL).
FT
SQ      SEQUENCE 447 AA; 49824 MW; 042BDD796ADD2A6 CRC64;

Alignment Scores:
Pred. No.:                               Length:           447
Score:                                   Matches:            46
Percent Similarity:                     37.10%
Best Local Similarity:                   20.81%
Query Match:                             Mismatches:        75
                                           Indels:              64
                                           Gaps:                9

US-09-830-244B-2 (1-1331) x TBBL_CYPAP (1-447)
QY      365 GTGGGAAGATGTCCCATTTACCATTCACAGTCTTTTGATGTGCACATAATGCATTATTGGT   424
Db      152 llaalylsvalatrgluciluntyrProaspalgmetwctcstfhyrServalphepro   171
QY      425 AATGCCGTTATGTGGGAATACAGCATAGTAGTAATAACTGTTAACAGTAATAACTRACAC---   481
Db      172 serProLyvalSerAsphrThyValValGluprotTyraNaLatnrLeuSerValIHISglN   191
QY      482 -----TTGGATTGGTCGACCGCTCTGCACATATACCT   511
Db      192 leuValIGluasnAlaaspcluvaiMetValIIeaSpnsgnuIalaLeuTYrAspIIecys   211

```

```

OY 512 TTTGAATGACGAAGAGCTTTAAAGAGAAA-----GAGGCA 54
    ||| |||||::: |||
Db 212 PheargpnrleuLysleuThrThrProThrTyrgLysleuAsnHisleuValSerIle 23
    ||| |||||::: |||
OY 548 TGTCTGCAGAAAGAGATAGCTAATATTTTTTGTACTTATCTGAATCCAAAGATGCTGC 60
    |||:::
Db 232 CysIleSer-----GlyValThrcysCys 23
    -TTCCCTCGAGTGTCTTTTCCCTCTTACGATCC-----TCATTTGAATCCC 65
OY 608 |||||::: ||| ||||| |||||
Db 240 LeuargpnrProGlyGlnLeuAsnSerAspLeuArgLysleuAlaValAsnLeuIlePro 25
    |||||::: ||| |||||::: |||
OY 653 CTCTGGAGACACAGACAGTATAGTAGAACTCTCCATTTCTTGTGTGTTTTT----- 70
    |||||::: ||| |||||::: |||
Db 260 Phe-----ProargLeuHisPheMetIleGlyPheValPro 27
    |||||::: ||| |||||::: |||
OY 707 AAGACAGAGACTCGTCTCAAAAAAAGACATTATCATTAACATCTTATT----- 76
    ||| |||||::: ||| |||||::: |||
Db 273 LeuThSerargLysSerGlnGlnTyrrArgAlaLeuThrValProGlnLeuThrGlnGln 29
    ||| |||||::: ||| |||||::: |||
OY 761 -----AGAGCCCTATTTCTTATCT 78
    |||
Db 293 MetPheaspAlaLysAsnMetCysAlaAlaAspProArgHisGlyArgTyrrLeuThr 31
    ||| |||||::: ||| |||||::: |||
OY 782 GAGGCGACCTTTTTTTTTTTTAAACGTTAAGTACTGATGCTCAACAGACAAATATTTCTG 84
    ||| |||||::: ||| |||||::: |||
Db 313 AlSerAlaLeuPheargLysArgMetSerThrLysGlnValAspGlnGlnMetLeuAsn 33
    ||| |||||::: ||| |||||::: |||
OY 842 ATCAGATATGATCCCTCTCAACACTGACCAATGCGTTTCATTAAGCGGAGAAACAGC 90
    ||| |||||::: ||| |||||::: |||
Db 333 ValGln-----AsnLysAsnSerSerTyrrPheValGlnIrrPleProAsnAsn 34
    ||| |||||::: ||| |||||::: |||
OY 902 ATT 904
    |||
Db 349 Ile 349

RESULT 13
IL8B_PANTR
ID IL8B_PANTR STANDARD; PRT: 353. AA.
AC 028807;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE High affinity interleukin-8 receptor B (IL-8R B) (CXCR-2) (Fragment).
GN IL8R or CXCR2.
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.
OX NCBI_TaxId=9598;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96175151; PubMed=9110929;
RA Alvarez V., Coto E., Setien F., Gonzalez S., Gonzalez-Roces S.,
RA Lopez-Iarrea C.;
RT "Characterization of interleukin-8 receptors in non-human primates.";
RL Immunogenetics 43:261-267(1996).
CC -I- FUNCTION: RECEPTOR TO INTERLEUKIN-8, WHICH IS A POWERFUL
CC NEUTROPHILS CHEMOTACTIC FACTOR. BINDING OF IL-8 TO THE RECEPTOR
CC CAUSES ACTIVATION OF NEUTROPHILS. THIS RESPONSE IS MEDIATED VIA A
CC G-PROTEIN THAT ACTIVATE A PHOSPHATIDYLINOSITOL-CALCIUM SECOND
CC MESSENGER SYSTEM. THIS RECEPTOR BINDS TO IL-8 WITH A HIGH AFFINITY
CC AND TO GNG/G13A AND NRP-2 ALSO WITH A HIGH AFFINITY.
CC -I- SUBCELLULAR LOCATION: Integral membrane protein.
CC -I- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```

QY 735 CCTTTTGTGGAGCAGAGCTCTGCTTTAAATAACAAACAAAGAAATGAGAGTTCTA 678
Db 197 -----MetGlyValasnThrAlaSerTrpArgMetLeu 207
QY 675 CTAACTGTCCTGTCCTCCAGAGGGAGTTCATGAGAGT----- 637
Db 208 LeuArgMetLeuProGlnSerPheGlyPheIleValProIleLeuIleMetLeuPheCys 227
QY 636 -----CGTAAGAAAGAAACAAACACCTCCAGGGAGACAGCATTTGGATT 592
Db 228 TTTGCTPheThrLeuArgThrLeuPheLysAlaIleMetGlyIleLysHisArgAlaMet 247
QY 591 CAGATTAAGTACCAAAAAATATGATAGCTATCTTTTCCAGACATGCGCTTTCTGT 532
Db 248 ArgValIlePheMetAlaValIleLeuIlePheLeuLeuCysTrp-LeuProTyrHisLeuVal 267
QY 531 AACACTTTC-----GTCATTCAAAAGCGCATTTGGTAGA 498
Db 267 LeuLeuAlaAspThrLeuMetValGhrGlnValIleGlnGlnIutThrCysGlnArg 285
RESULT 14
TBB_TRYBR
AC P04107, STANDARD; PRT: 442 AA.
DT 01-NOV-1986 (Rel. 03, Created)
DT 01-NOV-1986 (Rel. 03, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Tubulin beta chain.
OS Trypanosoma brucei rhodesiense.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
RX NCBI_TaxID=31286;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=86006273; PubMed=4043732;
RA Kimmel B.E., Samson S., Wu J., Hirschberg R., Yairnough L.R.;
RT "Tubulin genes of the African trypanosome Trypanosoma brucei
RT rhodesiense: nucleotide sequence of a 3.7-kb fragment containing genes
RL for alpha and beta tubulins."
RL Gene 35:237-248(1985).
RN [2]
RP SEQUENCE OF 1-8 AND 433-442 FROM N.A.
RA MEDLINE=85298227; PubMed=2994042;
RA Sather S., Agabian N.;
RT "A 5' spliced leader is added in trans to both alpha- and
RT beta-tubulin transcripts in Trypanosoma brucei."
RL Proc. Natl. Acad. Sci. U.S.A. 82:5695-5699(1985).
CC -1- FUNCTION: TUBULIN IS THE MAJOR CONSTITUENT OF MICROTUBULES. IT
CC BINDS TWO MOLES OF GTP, ONE AT AN EXCHANGEABLE SITE ON THE BETA
CC CHAIN AND ONE AT A NONEXCHANGEABLE SITE ON THE ALPHA-CHAIN.
CC -1- SIMILARITY: BELONGS TO THE TUBULIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.slb.ch).
CC -----
DR EMBL: K02836; AAA30261.1; -
DR EMBL: M11747; AAA30265.1; -
DR EMBL: M11748; AAA30268.1; -
DR PIR: A02976; UBRUB.
DR InterPro: IPR002453; Beta_tubulin.
DR InterPro: IPR000217; Tubulin.
DR InterPro: IPR003008; Tubulin_FtsZ.
DR Pfam: PF00091; tubulin_1.
DR PRINTS: PR01161; TUBULIN.
DR PROSITE: PS00227; TUBULIN_1.
DR PROSITE: PS00228; TUBULIN_B_AUTOREG_1.
KM Microtubules: GTP-binding.
NT NP_BIND 140 146 GTP (POTENTIAL).

```

SQ SEQUENCE 442 AA: 49704 MW: 628040DB20170A4B CRC64;
 Alignment Scores:
 Pred. No.: 2.11 Length: 442
 Score: 83.00 Matches: 47
 Percent Similarity: 39.81% Conservative: 35
 Best Local Similarity: 22.82% Mismatches: 76
 Query Match: 3.57% Indels: 48
 Gaps: 8
 DB: 1
 US-09-830-244B-2 (1-1331) x TBB_TRYBR (1-442)
 QY 386 TTTTCAGATCTTTGATGTCACATTAATGCTATGCTTATGCTTATGCTGAATA 445
 Db 159 TyrProaspArgIleMetMetThPheSerIleIleLeuSerProlyValSerAspThr 178
 QY 446 AGCATAGTAAATGTAAGTAACTGTAAGTAACTACAC----- 481
 Db 179 ValValGluProGlyrAsnThrThrLeuSerValHisGlnLeuValGluAsnSerAspGlu 198
 QY 482 -----TTGATTTTCTGTCACCTCTACCATTAAGCTTTGATGCTGAAGTAACTGA 532
 Db 199 SerMetCysIleAspAsnGluAlaLeuThrAspIleCysPheArgThrLeuLysLeuThr 218
 QY 533 ACAGAGAAAGAGCATGTCTGCAGAAAGAGATAGTAAATATTTTGTGCTTATATCTGA 592
 Db 219 ThrProThrPheGlyAspLeuAsnHisLeuValSerAlaVal-----ValSer 234
 QY 593 AATCCAAAGATGCTGC-----TTCCCTGAGGCTTTTCTCTGCTTACGATCC----- 640
 Db 235 GlyValThrCysCysLeuArgPheProGlyGlnLeuAsnSerAspLeuArgLysLeuAla 254
 QY 641 ----TCAITGAAITCCCTCTGGAGACACAGACAGATAGTAACTCATTTCTTGTGT 697
 Db 255 ValAsnLeuValProPhe-----ProArgLeuHisPheHeMet 267
 QY 698 TTGTTTATTAAG-----ACAGAGACTGCTCTCAAAAAAGACATTATCATTAATA 751
 Db 268 MetGlyPheAlaProLeuThrSerArgLysGlnGlnTyrArgGlyLeuSerValPro 287
 QY 752 CATCTTATTT----- 766
 Db 288 GluLeuThrGlnGlnMetPheAspAlaLysAsnMetGlnAlaAlaAspProArgHis 307
 QY 767 CCTAATTTCTTATCTGAGGCGCTCTTTTATTAACAGTAAAGTACGATGTCAC 826
 Db 308 GlyArgTyrLeuThrAlaSerAlaLeuPheArgLysArgMetSerThrLysGluValAsp 327
 QY 827 AGACAATATTTCTGATCAGATAGTCCCTGCAACAGTACCAATGCTTATTAAG 886
 Db 328 GluGlnMetLeuAsnValGln-----AsnLysAsnSerSerTyrPheIleGlu 343
 QY 887 TGGGAGAGAAACACCATTT 904
 Db 344 TrpIleProAsnAsnIle 349
 RESULT 15
 ID N120_YEAST STANDARD: PRT: 1037 AA.
 AC P35729: P35730:
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Nucleoporin NUP120 (Nuclear pore protein NUP120).
 GN NUP120 OR RAT2 OR YKL057C OR YKL314 OR YKL313.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes;
 CC NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=8288C;
 RX MEDLINE=94378723; PubMed=8091862;

RA Rasmussen S.W.;
 RT "Sequence of a 28.6 kb region of yeast chromosome XI includes the
 RT FPA1 and TOA2 genes, an open reading frame (ORF) similar to a
 RT translationally controlled tumour protein, one ORF containing motifs
 RT also found in plant storage proteins and 13 ORFs with weak or no
 RT homology to known proteins.";
 RL Yeast 10:563-568(1994).
 RN (2)
 RP REVISIONS.
 RA Rasmussen S.W.;
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
 RN (3)
 RP CHARACTERIZATION, AND SEQUENCE OF 189-206 AND 800-807.
 RX MEDLINE=96134019; PubMed=8557736;
 RA Aitchison J.D., Blobel G., Rout M.P.;
 RT Nup120p: a yeast nucleoporin required for NPC distribution and mRNA
 RT transport.";
 RL J. Cell Biol. 131:1659-1676(1995).
 RN (4)
 RP PARTIAL SEQUENCE OF 550-555: 799-803 AND 375-384.
 RX MEDLINE=96152656; PubMed=8565072;
 RA Siniosoglou S., Wimmer C., Rieger M., Doye V., Tekotte H., Weise C.,
 RA Emig S., Segref A., Hurt E.C.;
 RT "A novel complex of nucleoporins, which includes Sec13p and a Sec13p
 RT homolog, is essential for normal nuclear pores.";
 RL Cell 84:265-275(1996).
 CC -1- FUNCTION: REQUIRED FOR EFFICIENT MRNA EXPORT FROM THE NUCLEUS TO
 CC -1- SUBUNIT: INTERACTS WITH NUP84, NUP85, SEC13 AND SEH1.
 CC -1- SUBCELLULAR LOCATION: Nuclear pore complex.
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb-sdb.ch).
 CC
 DR EMBL: X75781; CA53445.1; ALT_SEQ.
 DR EMBL: X75781; CA53444.1; ALT_SEQ.
 DR EMBL: Z28057; CA481894.1; -
 DR PIR: S37879; S37879.
 DR SGD: S0001540; NUP120.
 KW Nuclear protein; Transport; Coiled coil.
 FT DOMAIN 435 465
 FT SIMILAR 434 763 TO NUP133 (AA 98-420).
 FT CONFLICT 384 384 K -> R (IN REF. 4).
 SO SEQUENCE 1037 AA: 120447 MW: D4655E6116C54503 CRC64;
 Alignment Scores:
 Pred. No.: 2.34 Length: 1037
 Score: 83.00 Matches: 78
 Percent Similarity: 39.48% Conservative: 29
 Best Local Similarity: 28.78% Mismatches: 81
 Query Match: 3.57% Indels: 83
 DB: 1 Gaps: 19
 US-09-830-244B-2 (1-1331) x N120_YEAST (1-1037)
 QY 463 TGTATC-----AGTAATCTACACTTGATTTGCTGCACCTTACCAATAGCTTT 513
 Db 69 CysTyrHisPheSerArgSerThrLeuLeuThrPheTyrProLeuSerAspAlaTyr 88
 QY 514 -----TGAATGCTGAAGTGTATACACAGAGAAAGGCGATG 549
 Db 89 HisGlyLysThrIleAsnIleHisLeuProAsnAlaSerMetAsnGlnArgTyrThrLeu 108
 QY 550 TCTGCAGAAAGAGTACTAATATTTTGTGTA-----CTTATCTGAATCCAGATGCTG 606
 Db 109 ThrIleGlnValGlnGlnGlnLeuLeuValAsnValIleLeuLysAspGlySerPhe 128
 QY 607 CTTCCTCGAGGTG---TTTCCTTCTTACGATCTCATTAATCCCTC----- 655

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: June 24, 2003, 19:16:12 ; Search time 90.5 Seconds

(without alignments)
6060.745 Million cell updates/sec

Title: US-09-830-244B-2

Perfect score: 2322
Sequence: 1 ctatgttttagatcaaga.....gcctttccatgtaactgt 1331

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODE=frame+np.model -DEV=xlp
-Q/cgn2.1/USPTO.spool/US09830244/runat.24062003.130001.10042/app.query.fasta.1.1479
-DB=SPREMBL.21 -QFMT=fasta -SUFFIX=np.rspt -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdl
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NOR-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09830244 -CGCN.1.1.152 -runat.24062003.130001.10042 -NCPU=6 -ICPU=3
-NO_MAP -LANG=QUERY -NEG_SCORES=0 -WAIT -DSPELCK=100 -LONGLOG
-DEV_TIMESOUT=120 -WARN_TIMESOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

SPREMBL.21:
1: sp_archaea:
2: sp_bacteria:
3: sp_fungi:
4: sp_human:
5: sp_invertebrate:
6: sp_mammal:
7: sp_mhc:
8: sp_organelle:
9: sp_phage:
10: sp_plant:
11: sp_proteint:
12: sp_virus:
13: sp_vertebrate:
14: sp_unclassified:
15: sp_virus:
16: sp_bacteriophage:
17: sp_archaeophag:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	100.5	4.3	512 8 Q9XMS7	Q9XMS7 tetrahymena

2	96	4.1	387 5 Q9BKC6	Q9BKC6 malawimonas
3	92.5	4.0	673 10 Q9XG31	Q9XG31 guillardia
4	91.5	4.0	949 8 Q9BRX5	Q9BRX5 guillardia
5	91	3.9	387 5 Q9BKC6	Q9BKC6 malawimonas
6	91	3.9	2708 5 Q15791	Q15791 plasmodium
7	91	3.9	2742 5 Q15801	Q15801 plasmodium
8	91	3.9	2819 5 Q15792	Q15792 plasmodium
9	90	3.9	1560 5 Q9GRX5	Q9GRX5 dictyostell
10	89.5	3.9	307 10 Q9AVZ0	Q9AVZ0 guillardia
11	89	3.8	135 15 Q90814	Q90814 human immun
12	88.5	3.8	474 2 Q9ZHL1	Q9ZHL1 haemophilus
13	88.5	3.8	589 2 Q93Q09	Q93Q09 haemophilus
14	88	3.8	1005 10 Q98RM3	Q98RM3 guillardia
15	87.5	3.8	332 5 Q9XUN3	Q9XUN3 caenorhabd
16	87	3.8	227 5 Q8T1W4	Q8T1W4 dictyostell
17	87	3.8	855 10 Q9ZS83	Q9ZS83 lycopersico
18	87	3.8	1144 5 Q9G019	Q9G019 dictyostell
19	86.5	3.7	876 15 Q79670	Q79670 human immun
20	86	3.7	443 5 Q25316	Q25316 leishmania
21	86	3.7	2028 8 Q32064	Q32064 chlamydomon
22	85.5	3.7	627 12 Q9EMS5	Q9EMS5 amastax mo
23	85.5	3.7	3933 5 Q97239	Q97239 amastax mo
24	85	3.7	310 11 Q8VW4	Q8VW4 mus musculu
25	85	3.7	573 5 Q77334	Q77334 plasmodium
26	85	3.7	577 10 Q9SW08	Q9SW08 arabidopsi
27	85	3.7	620 8 Q9Z2P7	Q9Z2P7 cyanidiosi
28	85	3.7	672 5 Q963G8	Q963G8 plasmodium
29	85	3.7	2166 12 Q36635	Q36635 human respi
30	85	3.7	2391 5 Q27732	Q27732 plasmodium
31	85	3.7	3124 5 Q9GV88	Q9GV88 caenorhabd
32	84.5	3.7	311 5 Q966K6	Q966K6 caenorhabd
33	84.5	3.7	2313 8 Q8WHW9	Q8WHW9 psittacus nu
34	84	3.6	340 17 Q97A98	Q97A98 thermoplasm
35	84	3.6	429 12 Q9EN15	Q9EN15 amastax mo
36	84	3.6	651 10 Q9AW16	Q9AW16 guillardia
37	84	3.6	696 4 Q90U53	Q90U53 homo sapien
38	84	3.6	748 4 Q12311	Q12311 saccharomyc
39	84	3.6	1152 5 Q8T216	Q8T216 dictyostell
40	83.5	3.6	283 13 Q8UV46	Q8UV46 brachydanio
41	83.5	3.6	314 11 Q8VGY1	Q8VGY1 mus musculu
42	83.5	3.6	331 5 Q9U2Y0	Q9U2Y0 caenorhabd
43	83.5	3.6	314 5 Q9U2Y0	Q9U2Y0 caenorhabd
44	83.5	3.6	515 17 Q97W1	Q97W1 sulfolobus
45	83.5	3.6	1844 5 Q97287	Q97287 plasmodium

ALIGNMENTS

RESULT 1

ID	Q9XMS7	PRELIMINARY:	PRT:	512 AA.
AC	Q9XMS7			
DT	01-NOV-1999 (TREMBLrel. 12, Created)			
DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	Haem lyase.			
GN	YEJR.			
OS	Tetrahymena pyriformis.			
OG	Mitochondrion.			
OC	Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;			
OC	Tetrahymena; Tetrahymena.			
OX	NCBI_TaxID=5908;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20181866; PubMed=10715208;			
RA	Edgelist J., Burger G., Gray M.N.;			
RT	"Expression of mitochondrial protein-coding genes in Tetrahymena			
RT	pyriformis.";			
RL	J. Mol. Biol. 297:381-393(2000).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20181865; PubMed=10715207;			
RA	Burger G., Zhu Y., Littlejohn T.G., Greenwood S.J., Schnare M.N.,			

947 CTCGTGTCCTCTTAGAAAAAAACCTCCACCTTACGTCTCATTTATACCTC 1006
 333 -----asn-vallysalaalavalcysaspileproply 344
 1007 TACTTCCAAGTTAATATTCTATTCTCGATGATGCTTTATACCAAGAAGCCTTATCA 1066
 |||:::|||||


```

Db      692 LeuValIleLeuLeuLysAsnPhelIleSerIleLeuPheLeuThr----- 706
QY      1085 TGGTGTGTCTGGAAGCTGCTTAAGGCTTTGGTATTAAGCAATATCCAGAAATAG 1026
Db      707 -----TleLysTyrAspLysLysIleLeuAsnSerAsnLys 719
QY      1025 ATAAATTAACCTTGG-----AACTAAGGAGATTAATATGACACAGTAAGGCTGAGGCTT 970
Db      720 ---TleAsnPhetylLysAsnPhelIleLysValSerPheIleLeuLysAsnPh 738
QY      969 TTTTCTTAAAGGAGACAGAGT-----TTTATATTAATCAATC 928
Db      739 PheGlnIleLeuAsnPhLeuSerIlePheLeuTyrIleLeuPheIleLeuAlaGlyIle 758
QY      927 AGTCATCCCAAAAGTAC---TTTAAATGCTGTTTCTTCC-----CACTTATG 880
Db      759 IleserSerLysSerPheLysIleLysLeuIlePheAsnSerIleIleLeuGlyPheLys 778
QY      879 AAACCAATTTGCTACTGTTGACAGGGGAC-----TATCTGATCAGAAATATTGT 829
Db      779 LysPheLysSerPheSerLeuAspSerLysLeuAlaGlyAlaPhePheLysAsnPhLeu 798
QY      828 CTGTTGACATCAGTACTTAACTGTTTAAATAAAACAGTGCCTTCAGATAAGAAATTA 769
Db      799 ILeValSerSerIleIle-----LeuLysLysSerIleLeuSerIlePheLys--- 815
QY      768 GGGGCTCTAATAGATGTATTAATGATTAATGTCCTTTTTCAGACAGAGTCTGTC 709
Db      815 ----- 815
QY      708 TTAATAAAACAACAAAGAAATGAGAGTCTACTACTGCTCTGCTCCAGAGGGA 649
Db      816 ---LysSerLeuPheLysAspPheLeuIleLeuLeuThrGlnIle----- 829
QY      648 TTCATAGAGATGCTGAAGAAGAAAC-----AACTGACAGGGAGACGACT 601
Db      830 ILeIleLeuLysArgLysLysGlyLysPheTyrValLeuAsnLeuIleAsnLeuThrPhe 849
QY      600 CTGGGATTCAGATTAAGTAAAGTACCAAAATATTAAGTATCTCTTTCGACAGATGCTCT 541
Db      850 LeuAlaLeuAsnSerAsnGlnLysAsnIlePheLysLysPheIleLysLysIle 869
QY      540 TTCTGTTAAACACTTCACTGCTCAAAAGGCTATGCTAGAGGTCAGCAATCCAG 481
Db      870 PheGlyIle-MetIle---TleIleHisArgHisIleValLysIleLeuLysAsnSerAs 888
QY      480 TGTAGATTTACTGTAACGTTTATTAATCACTATGCTGATTCACCAATAAGGCATTACCA 421
Db      888 nleuAspLeuLeuIleTyrLeuLysArgMetCysIleGlu----- 902
QY      420 ATAAAGGCAATTAGTGCACATCAAAAGATCTGAATAATGCTAATGGAGCATTTCCCAAAA 361
Db      903 -----LeuThrSerLeuGlyValAsnTyrLeuLysAsnHisTyrAsnGlnAs 918
QY      360 GGAAGAATCCCAATGGCTCTATAAGGAGCAATATAGC 321
Db      918 nAsnLysAsn-----LeuLysAlaIleTyrSer 927

RESULT 5
Q9BKCS PRELIMINARY: PRT: 387 AA.
AC 09BKCS:
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE Beta-tubulin (Fragment).
OS Malawimonas jakobiformis.
OC Eukaryota; Malawimonadidae; Malawimonas.
NCBI_TaxID=136089;
NC 11
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC50310;
RX MEDLINE=21165323; PubMed=11264402;

```

```

RA Edgcomb V.P., Roger A.J., Simpson A.G.B., Kysela D.T., Sogin M.L.:
RT "Evolutionary relationships among 'jakobid' flagellates as indicated
by alpha- and beta-tubulin phylogenies."
RL Mol. Biol. Evol. 18:514-522(2001).
DR EMBL: AF267186; AKS7437.1; -.
DR InterPro: IPR000217; Tubulin.
DR InterPro: IPR003008; Tubulin_FtsZ.
DR Pfam: PF00091; tubulin_1.
DR PRINTS: PR01161; TUBULIN.
DR PROSITE: PS00227; TUBULIN; UNKNOWN_1.
KW GTP-binding.
FT NON_TER 1
FT TER 387
SQ SEQUENCE 387 AA; 42840 MW; 6E3A6AC159AF90 CRC64;

Alignment Scores:
Pred. No.: 0.612 Length: 387
Score: 91.00 Matches: 44
Percent Similarity: 41.26% Conservative: 41
Best Local Similarity: 21.36% Mismatches: 73
Query Match: 3.92% Indels: 48
DB: 5 Gaps: 8

US-09-830-244B-2 (1-1331) x Q9BKCS (1-387)
QY 386 TTTTCAGATCTTTGATGTCGACTAATGCCATTAATGCGTATGCGCTATTGCGATAC 445
Db 144 TyrProAspArgMetCysThrPheSerValValProSerProLysValSerAspThr 163
QY 446 AGCATAGTAAATTAACGTGTACAGTAATCTACAC----- 481
Db 164 ValValGluProTyrAsnAlaThrLeuSerValHisGlnLeuValGluAsnAlaAspGlu 183
QY 482 -----TTGATTTGCTGCACCTCTACCAATAGCGCTTGAATGACGGAAGCTTA 532
Db 184 ValPheCysValAspAsnGlnAlaLeuTyrAspIleCysPheArgThrLeuLysLeuThr 203
QY 533 ACAGAGAAGAGGAGCATGCTGCAGAAAAGAGATGATAATTTTGGTACTTTATCTGA 592
Db 204 ThrProThrTyrGlyAspLeuAsnHisLeuValSerAlaVal-----MetSer 219
QY 593 AATCCAGATGCTGCG-----TTCCTGCGAGGTTGTTTCTTACGATCC----- 640
Db 220 GlyValThrCysCysLeuAlaGlyPheProGlyGlnLeuAsnSerAspLeuArgLysLeuAla 239
QY 641 ---TCATTTGAATCCCTCTGGGAGCAGACAGACAGTATGAGAATCTCCATTTCTTCTT 697
Db 240 ValAsnLeuIleProPhe-----ProArgLeuHisPhePheMet 252
QY 698 TTGTTTTTTTAAAG-----ACAGAGACTGTCTCAAAAAAAGACATTATCATTTATTA 751
Db 253 ValGlyPheAlaProLeuThrSerArgGlySerGlnGlnTyrArgAlaLeuThrValPro 272
QY 752 CATCTTATTT-----AGAGCC 766
Db 273 GluLeuValGlnGlnMetPheAspAlaLysAsnMetMetAlaAlaSerAspProArgHis 292
QY 767 CCTAATTTCTTATCTGAAGGACACTGTTTTTTTAAACAGTAAGTACGATGCTCAAC 826
Db 293 GlyHisTyrLeuThrAlaAlaValAlaPheArgGlyArgMetSerThrLysGlnValAsp 312
QY 827 AGACAAATATTTCTGTCGATGATAGTCCCTGTCACACAGTACAGTAATGCTTCAATAAG 886
Db 313 GluGlnMetLeuAsnValGln-----AsnLysAsnSerSerTyrPheValGlu 328
QY 887 TGGGAAGAAACAGCATTT 904
Db 329 TrpIleProAsnAsnVal 334

RESULT 6
ID 015791 PRELIMINARY: PRT: 2708 AA.
AC 015791;

```


DB 687 sSergIuTySerTyLysThrGlyValIleAsnIleHisAspLeuLeuAspTyrPh 707
 QY 1032 TCTGATATGCTTTATATCAAGAGCCTTATCAGCCAG----- 1071
 DB 707 ePhaspleuHisPheIleLysAspGluLeuLysCysGlnAspThrAsnGlyAsnIleLys 727
 QY 1072 -----TTCAGAACACCACTATACGACACCAACCAATACCAATATAT 1124
 DB 727 sPhaspgImetAsnAspHisValLysAsnAspHisThrAsnTyTyTyGlnAsnAsnAs 747
 QY 1125 AGCTTTATATCATCAGT 1143
 DB 747 nIlePheAsnLeuLeuThr 753

RESULT 8
 015792 PRELIMINARY; PRT; 2819 AA.

AC 015792
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE CG2.
 GN CG2.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-HB3;
 RX MEDLINE=98054002; PubMed=9393853;
 RA Su X., Kirkman L.A., Fujioke H., Wellens T.E.;
 RT "Complex polymorphisms in an approximately kDa protein are linked to
 chloroquine-resistant P. falciparum in Southeast Asia and Africa.";
 RL Cell 91:593-603(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-HB3;
 RX MEDLINE=20446201; PubMed=10989140;
 RA Fiodock D.A., Nomura T., Cooper R.A., Su X.-z., Talley A.K.,
 RA Wellens T.E.;
 RT "Allelic modifications of the cq2 and cgl genes do not alter the
 chloroquine response of drug-resistant Plasmodium falciparum.";
 RL Mol. Biochem. Parasitol. 110:1-10(2000).
 DR EMBL: AF030693; AAC47854.1;
 DR SEQUENCE 2819 AA; 335122 MW; FBF3CCCCOAC78B9D CRC64;

Alignment Scores:

Pred. No.: 0.669 Length: 2819
 Score: 91.00 Matches: 41
 Percent Similarity: 43.85% Conservative: 41
 Best Local Similarity: 21.93% Mismatches: 58
 Query Match: 3.92% Indels: 47
 DB: 5 Gaps: 9

US-09-830-244b-2 (1-1331) x 015792 (1-2819)

QY 669 CACTTGTAGTACACTCTTCATTTCTTTGTTTAAAGACAGACTGTGTCAAA 728
 DB 580 GAAValAlpHeMetLeuProPheSerTyTyPheSerAsnLysLysLys-----Lys 596
 QY 729 AAAAGACACTTATCATTTATACATCATTTATAGACCCCTAATTTCTTATGTAG--- 785
 DB 597 LysLys-----AsnIleIlePheLysAsnIleLeuSerTyTyTyGluTy 611
 QY 786 GCACGTGTTTTTTTAAACAGTAACTA-----CTGATGTCAACAGCAATATATTTC 839
 DB 612 ValIleCysPhePheMetLysLysLysIleLysAsnIleLeuSerTyTyAsnProTyTyTy 631
 QY 840 TGATGATAGTACCTGTGCAACAGTACCAATGTGGTTCTTAAGTGGGAAGAAACA 899
 DB 632 -IleHisIle-----GluSerTyTyTyAsnLysLysIleGly----- 642

QY 900 GCATTTAAAGTACTTTTGGGAGACTGATTG-----AGTAATAA 941
 DB 643 ---TYCysIleValPhePheLeuPheSerIleLeuLysIlePheCysThrCysAspAs 661
 QY 942 TAAAGCTGTGCTGCTTATACCAAGAGCCTTATCAGCCAGT----- 984
 DB 661 PASpThrThrLeuLeuIleSerTyTyGlyShiValTyTyTyLeuSerAspSerLeuGly 681
 QY 985 -----TACTGTGCTATTTATATCCCTTAGTCCCAAGTTAATATCTTAT 1031
 DB 681 sSergIuTySerTyLysThrGlyValIleAsnIleHisAspLeuLeuAspTyrPh 701
 QY 1032 TCTGATATGCTTTTATATCAAGAGCCTTATCAGCCAG----- 1071
 DB 701 ePhaspleuHisPheIleLysAspGluLeuLysCysGlnAspThrAsnGlyAsnIleLys 721
 QY 1072 -----TTCAGAACACCACTATACGACACCAACCAATACCAATATAT 1124
 DB 721 sPhaspgImetAsnAspHisValLysAsnAspHisThrAsnTyTyTyGlnAsnAsnAs 741
 QY 1125 AGCTTTATATCATCAGT 1143
 DB 741 nIlePheAsnLeuLeuThr 747

RESULT 9

09GRX5 PRELIMINARY; PRT; 1560 AA.

AC 09GRX5
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Hypothetical 175.1 kDa protein.
 GN GDT1.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyostelidia; Dictyostelium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AX2;
 RA Primpke G., Salger K., Wetterauer B.W.;
 RT "Absence of complementation in nonallelic mutants of Dictyostelium
 RT discoideum with defects in the transition from growth to
 RT development.";
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AX2;
 RX MEDLINE=20255223; PubMed=10793140;
 RA Zeng C., Anjard C., Riemann K., Konzok A., Nellen W.;
 RT "gdt1, a new signal transduction component for negative regulation of
 RT the growth-differentiation transition.";
 RL Mol. Biol. Cell 11:1631-1643(2000).
 DR EMBL: AJ279060; CAC09932.1;
 DR InterPro: IPR000719; Euk_Pkinase.
 DR InterPro: IPR002290; Ser_Tyr_kinase.
 DR Pfam: PF00069; Pkinase.
 DR ProDom: PD000001; Euk_Pkinase; 1.
 DR SMART: SM00220; S_TKc; 1.
 DR SMART: SM00219; TyrKc; 1.
 DR PROSITE: PSS0011; PROTEIN_KINASE_DOM; 1.
 DR ATP-binding: Hypothetical protein; Transferase.
 FT CONFLICT 20 20 I -> V (IN REF. 2).
 FT CONFLICT 619 619 F -> S (IN REF. 2).
 SO SEQUENCE 1560 AA; 175089 MW; 883B5BF9AD9E81EA CRC64;

Alignment Scores:

Pred. No.: 0.838 Length: 1560
 Score: 90.00 Matches: 56
 Percent Similarity: 39.27% Conservative: 41
 Best Local Similarity: 22.67% Mismatches: 106
 Query Match: 3.88% Indels: 44
 DB: 5 Gaps: 7

US-09-830-244B-2 (1-1331) x Q9GRX5 (1-1560)

QY 577 TTGGTACTTATCTGAATCCAGATGCTGCTCCCTG-----CAGGTT 621

DB 1316 ILEISERATGLeuLysHisGlnSerIleLeuProLeuIleGlyCysValIleSerLys 1335

QY 622 GTTTCCTTCTTACGACCTCATTTGAATCCCTCGGAGCAGACAGCATTAAGTAAC 681

DB 1336 AspTyrLeuCysLeuAlaPheGluTyrProPoleuGlySerLeuAspTyr 1352

QY 682 TCTCCATTTCTTTGTTTGTTTTAAAGACAGACATCTCTCAAAAAGGACATT 741

DB 1353 -----IleIleSerLysLysLysLys 1360

QY 742 ATCATATACATCTTATTAGACCCCTTAATTTCTTAATCTGAAGCAGCTGTTT 801

DB 1361 MetSerIleThrGlnLysIleArgIleLeuIleAspValAlaLysGlyCysLysPheLeu 1380

QY 802 TAAACAGTTAAGTACTGATCAACAGACAAATATTGATTCAGATGCTCCCTGTC 861

DB 1381 -----GlnGlnSerIleIleGlnLysThrLeuArgAlaArg 1393

QY 862 CAGTAGCAATGCTGTTTCATTAAGTGGGAGAAACAGCATTTTAAAGTATTG 921

DB 1394 AsnIlePheLeuTyr-----AspThrAsnGlnAsnAlaGluValCysAlaLys 1409

QY 922 GAGACGATTGAGATTAATTAACCTGCTGCCCTTAAGAAAAAACCTTCCAC 981

DB 1410 ValLeuAspLeuThrSerSerLysThrIleLysGlyLeuAla----- 1423

QY 982 CTTCAGTGTCTATATATATCCCTTAGTCCAAAGTTATTTATCTTATCTGATAT 1041

DB 1424 -----CysAsnAsnTyrIleGluArgValAspThrProIleAsnLeuThrArgGluIle 1441

QY 1042 GCTTTTATA-----CCAAAGAGCCTTATACGACGATCCAGAACCAACACTATACGCAC 1095

DB 1442 SerIleIleArgAspProLysGlnAsnAsnAspPheAsnAsnSerAsnAsnSerAsn 1461

QY 1096 AACCATACCAACACATACCAATATAGCTTTATATCATCATTAACATGAGAGACA 1155

DB 1462 AsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnSerAsn 1481

QY 1156 TGATTTATGAGCTTGTGCGCAATATGACCTTCATCATATTTCTTCTTAC 1215

DB 1482 AsnSerSerSerLeuTyrAsnIle-----HisSerPheAlaValIleSerTyr 1498

QY 1216 CATATCACTACTACCACTTTTGTNAGATCATCTAAGAGCAATGCGAATGAACCC 1275

DB 1499 GluLeuLeuIleAspGluIleLeuValGlyAspThrArgLysPheGlyGlnGluLysPro 1518

QY 1276 TATTAATTTACTGATCTCTT 1296

DB 1519 SerIleGlyLeuAspLysIle 1525

RESULT 10

Q9AVZ0 PRELIMINARY; PRT; 307 AA.

AC Q9AVZ0;

DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE Hypothetical 37.7 kDa protein.

OS Guillardia theta (Cryptomonas phl.).

OC Eukaryota; Cryptophyta; Cryptomonadaceae; Guillardia.

OX NCBI_TaxID=55529;

IN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20087226; PubMed=10618395;

RA Zauner S., Fraunholz M., Mastl J., Penny S.L., Beaton M.,

RA Cavalier-Smith T., Maier U., Douglas S.,

RT "Chloroplast protein and centosomal genes: a tRNA intron, and odd

RT telomeres in an unusually compact eukaryotic genome, the cryptomonad

RT nucleomorph";

RL Proc. Natl. Acad. Sci. U.S.A. 97:200-205(2000).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=21223349; PubMed=11323671;

RA Douglas S., Zauner S., Fraunholz M., Beaton M., Penny S., Deng L.,

RA Wu X., Reith M., Cavalier-Smith T., Maier U.;

RT "The highly reduced genome of an enslaved algal nucleus.";

RL Nature 410:1091-1096(2001).

DR EMBL: AJ010592; CAC27081.1;

KW Hypothetical protein.

SQ SEQUENCE 307 AA; 37661 MW; 37E28BF8A1E0151D CRC64;

Alignment Scores:

Pred. No.:	0.883	Length:	307
Score:	89.50	Matches:	39
Percent Similarity:	39.54%	Conservative:	28
Best Local Similarity:	23.08%	Mismatches:	47
Query Match:	3.87%	Indels:	55
DB:	10	Gaps:	8

US-09-830-244B-2 (1-1331) x Q9AVZ0 (1-307)

QY 987 GTAAGGTGAGAGGCTTTTCTTAAAGGAGACAGACTTTATTATTCTCAATTC 928

DB 48 IleLysGlnLysLysPhePheLeuGluLeuAsp----- 59

QY 927 AGTCTCCAAAGTACTTTAAATGCTGTTTCTTCCACTTATGAACACCATTTG 868

DB 60 -----GlyTyrIleAsnAsnIlePheAsp----- 67

QY 867 CTACTGTTGACAGGGAGCTATCTGATCAGAAATATTTGCTGTGACATCAGTACTAAC 808

DB 68 TyrValLeuLeuSerGluTyrLeuIlePheAsnIleAsnLeuLysAsnAsnLeuIleSer 87

QY 807 TGTTTAAAAAAGAACAGTGCCTTCAGATAGCAATTAAGGCGCTTAATTAAGATGTTAT 748

DB 88 -----LysLysIlePheLeuLeuGluLysTyr 97

QY 747 AATGATAAATGCTCTTTTGTGAGACAGACAGCTGCTGTTAAAC----- 700

DB 98 -----TyrProGluPheAsnLysIleIleLeuLeuIleValAsnLeuPheValTyr 114

QY 699 -----AAAACAAAGAAATGAGAGATTCTAATCTGCTGCTCCAGAG 652

DB 115 AspAsnSerPheLysArgLysPheTyrLysPhePheIleSerLysPheCysThrTyrArg 134

QY 651 GGAATC-----AATGAGATGCTAAGAGAAAC 622

DB 135 AsnIleIleIleThrPheLeuAspIleLysValSerLysIleGluLysValPheAsn 154

QY 621 AACCTGACAG---GGAGAGCAGCATCTTGATTTACATTAAGCAAAAATTAAGCT 565

DB 155 SerIleLysThrAsnLysAsnTyrPheAsnPheGluIleLysLeuArgLysIlePheMet 174

QY 564 ATCTCTTTCTGC-----AGACAT 547

DB 175 IleLeuPheCysThrPhePheArgHis 183

RESULT 11

Q90814 PRELIMINARY; PRT; 135 AA.

AC Q90814;

DT 01-NOV-1998 (TREMBLrel. 08, Created)

DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE Envelope protein (Fragment).

OS Human immunodeficiency virus type 2.

OC Viruses; Retroid viruses; Retroviridae; Lentivirus.

OX NCBI_TaxID=11709;

IN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-A:
 MEDLINE=97255645; PubMed=910092;
 RA Xiang Z., Ariyoshi K., Wilkins A., Dias F., Whittle H., Breuer J.,
 RT "HIV type 2 pathogenicity is not related to subtype in rural guinea
 bissau".
 RL AIDS Res. Hum. Retroviruses 13:501-505(1997).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN-A:
 MEDLINE=98406190; PubMed=9733826;
 RA Grassly N., Xiang Z., Ariyoshi K., Aaby P., Jensen H., Dias F.,
 RT Van der Loeff, Whittle H., Breuer J.,
 RT "Mortality among human immunodeficiency virus type 2-positive
 villages in rural guinea-bissau is correlated with viral genotype".
 RL J. Virol. 72:7895-7899(1998).
 DR EMBL: AJ011256; CA009567.1;
 DR InterPro: IPR000777; GP120.
 DR Pfam: PF00516; GP120.1.
 KM AIDS: Coat protein; Glycoprotein.
 FT NON_TER 1 135
 FT SEQUENCE 135 AA; 15598 MW; 2A732E706BBE638 CRC64;

Alignment Scores:
 Pred. No.: 0.965 Length: 135
 Score: 89.00 Matches: 31
 Percent Similarity: 39.34% Conservative: 17
 Best Local Similarity: 25.41% Mismatches: 38
 Query Match: 3.83% Indels: 36
 DB: 15 Gaps: 5
 US-09-830-244B-2 (1-1331) x 090814 (1-135)

QY 916 TTTTGGAGAGCTGATTGAGTAATATAAAGCTGCTGCCCTTAGAAAAAACC 975
 DB : : : : :
 DB 15 TTTTTPHSSer-----LysAsnAsnLysThrIleSerLeuAsnLysTyrTyrAsn 32
 QY 976 TTCACCTTACTGT-----GTCATTATATCCCTTAGTTCACA 1014
 DB : : : : :
 DB 33 LeuThrLeuHisCysLysArgProGluAsnLysThrValAlaProIleThrLeuMetSer 52
 QY 1015 AAGTTAATATCTATATTCGATATTCGTTTATACCAAGACCTTATACACCGCTTC 1074
 DB : : : : :
 DB 53 GlyLeuValAlaPheHisSerGlnProIleAsnThrLysProArgGlnAlaGlnCysGlnPhe 72
 QY 1075 CAGAAC-----AACCATATAG 1092
 DB : : : : :
 DB 73 LysGlyLysGlnLysLysAlaMetLeuGluValLysGlnThrLeuValLysHisProLys 92
 QY 1093 CACAACCATACCAACCAATACCAATATACGTTTAT----- 1134
 DB : : : : :
 DB 93 TyrLysGlnThrAsn--AspThrAsnAsnIleHisPheThrAlaProGluLysGlySer 111
 QY 1135 -----ATCATCAGTAAGTGCAGCATGATATATGAGCTGCTG 1176
 DB : : : : :
 DB 112 AsnProLysValAlaTyrMetLeuThrAsnCysArgGlyLysPheLeuSerLeuGlnTyr 131
 QY 1177 CAATA 1182
 DB : : : : :
 DB 132 Aspleu 133
 RESULT 12
 092H1 PRELIMINARY: PRT: 474 AA.
 AC 092H1:
 DT 01-MAY-1999 (TREMblrel. 10, Created)
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
 DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
 DE LSPB (Fragment).
 GN LSPB.
 OS Haemophilus ducreyi.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.

OX NCBI TaxID=730;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=35000.
 RX MEDLINE=99030326; PubMed=9811662;
 RA Ward C.K., Lumbley S.R., Latimer J.L., Cope L.D., Hansen E.J.,
 RT "Haemophilus ducreyi secretes a filamentous hemagglutinin-like
 protein".
 RL J. Bacteriol. 180:6013-6022(1998).
 DR EMBL: AF057696; AAC79760.1;
 FT NON_TER 1 474
 FT SEQUENCE 474 AA; 53579 MW; 8796871FD446A213 CRC64;

Alignment Scores:
 Pred. No.: 1.16 Length: 474
 Score: 88.50 Matches: 42
 Percent Similarity: 38.46% Conservative: 33
 Best Local Similarity: 21.54% Mismatches: 71
 Query Match: 3.83% Indels: 49
 DB: 2 Gaps: 10
 US-09-830-244B-2 (1-1331) x 092H1 (1-474)

QY 885 TTTTGAACACCATTTGCTACTGTTGACAGGGAC---TATCTGATCAGAAATTTGT 829
 DB : : : : :
 DB 82 PheValLeuProAsn-----LeuSerGlyLysIlePheSerIleHisAspleuAsp 98
 QY 828 CTGTTGACATCAGTACTTAACTGTTTAAAAAAAACA-----CTGCCCTTACAT 778
 DB : : : : :
 DB 99 GlnLeuValGluValLeuAsnThrValAsnLysArgAlaGluIleLysValLeuAlaSer 118
 QY 777 AAGAAATTAAGGGCTCTATATAAGATGTTATATGATAATGCTTTTTCGACAGA 718
 DB : : : : :
 DB 119 LysAlaTyrGlyLys-----SerAsn 125
 QY 717 GTCCTGCTTAAAAAACAAGAAAGATGAGAGCTTCTACTTACTGCTGCTGCTC 658
 DB : : : : :
 DB 126 LeuAsnIleLeuThrGlnArgThrArgLysThrProThr--ValThrLeuSerIleAsn 144
 QY 657 CAGAGGGATTCATAGAGATGCTAGAGAGAAACAACCTGCAGGGAGAGACATCTT 598
 DB : : : : :
 DB 145 AsnSerGly-----LysLusAsnGlnGlnAsnGlyArgAsnGlnMet 158
 QY 597 GGATTTCGATTAAGTACCAAAATATTTAGCTATCTCTTTCGACAGATGCTTTC 538
 DB : : : : :
 DB 159 ThrLeuAsnValSerThrSerAspLeuLeuGlyThrAsn----- 171
 QY 537 TCTGTTAAGCTTCACTGATTCATTAAGGCTATGTAAGCTGCAGCAATCCACTGT 478
 DB : : : : :
 DB 172 -----AspValIlePheSerPheLysThrGlyTyr-----ArgLeuTyrLysGluThrLys 187
 QY 477 AGATTTCGTAACAGGTTTATTTAATCTATGCTGATACCAATTAAGCGATTAACCAATA 418
 DB : : : : :
 DB 188 LysAsnThrGlnGlnAsnTyrSerLeuSerTyrIleGlnProPheSerTyrTyrThrLeu 207
 QY 417 ATGCATTAGTCCATCAAAAGATCTGAAGAATGCTAATGG----- 376
 DB : : : : :
 DB 208 -----GluIleLysAlaSerGlnSerAlaTyrAsnLysGluLeuThrGlyPhe 223
 QY 375 ---ACATTCCCAAGAAAGCAATCAATGCGCTCATATAA 334
 DB : : : : :
 DB 224 TyrThrTyrProSerSerGlyLysThrGlnThrAlaAsnIleLys 238
 RESULT 13
 093009 PRELIMINARY: PRT: 589 AA.
 AC 093009:
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE Large superantigen protein exporter.
 GN LSPB.
 OS Haemophilus ducreyi.

DB 333 snvalalavallysarlgileysvallyls-----LyslysserL 347
 QY 357 AAGAAATCCAAATGGCTCCATATAAGGAGAAATATGCAATGCAAGCAATGAAATTT 298
 DB 347 yslleuiletyrlylsleserlysglnlleseerilethrphelystyrrhelsylei 367
 QY 297 CTGTCAAAAGAAAGTACTTCTCAAGCAATCAGAAAAAGCTGTCCAAATTAAGCCCG 238
 DB 367 laphelylslylsyrrhelsylei-----GlnaslylslelleuileuSerleuphellei 384
 QY 237 GGAGCCACATTAAGC---CAGTATCTCAGATACGTATTTTGT---AGCCT 190
 DB 384 leleuemetilleserlyshistyrile-Asplys-ilevalleuiletyrlylsyasaSerle 403
 QY 189 TAGGGCCCTCAAT-----ATTCGCCCTCAT-----CTGCTAAACTTCACTT 148
 DB 403 uleucllytyrlysnllelleleasnlnhlsglyvalarglyvalleuasn---Asnle 422
 QY 147 GTTACTCACAAGTCTGATTAATTCAGCCCAAGTAACTT-----AACATTC 100
 DB 422 userlysrProleuaspserlystyrrargansSeranslletyrlysplyscyshtileva 442
 QY 99 CTATGGCAGTCAAAATTCAGACATTTGTAACACACTGTAATTTACGTTGGTACATG 40
 DB 442 lpheserlystyrrhelsylei-----Asntyrpheserlystyrrhelsylei 457
 QY 39 AGACAGCTTACCG 27
 DB 458 -----Serleupro 460

RESULT 15

ID 09XUN3 PRELIMINARY; PRT; 332 AA.
 AC 09XUN3;
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE T20B3.5, protein.
 GN T20B3.5.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 NC NCBI_TaxID=6239;
 RN 11
 RP SEQUENCE FROM N.A.
 RA Mortimore B.J.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RN 12
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9069613; Pubmed=9851916;
 RA none;
 RT "Genome sequence of the nematode C. elegans: A platform for
 RT investigating biology."
 RL Science 282:2012-2018(1998).
 DR EMBL: Z81593; CAB04743.1; ..
 DR InterPro: IPR003003; 7TM_chemo2.
 DR InterPro: IPR000168; 7TM_nematode.
 DR Pfam: PF01604; 7TM_5; 1.
 SO SEQUENCE 332 AA; 38162 MW; 5D903258D5357D9F CRC64;

Alignment Scores:

Pred. No.: 1.47 Length: 332
 Score: 87.50 Matches: 64
 Percent Similarity: 33.99% Conservative: 40
 Best Local Similarity: 20.92% Mismatches: 87
 Query Match: 3.77% Indels: 115
 DB: 5 Gaps: 15

US-09-830-244b-2 (1-1331) x 09XUN3 (1-332)

QY 529 GTTACAGAGAGAAAGCATGTCTGCAAGAAAGATAGCTAATATTTTGG----- 579
 DB 89 Valleuylsglnllelylsarlgileysvallyls-----Tyrphelyvalleuval 107

QY 580 -----GTACTTAT 588
 DB 108 glylatyrmetvelalalleseerilevalphegluasnargleuphevalleuval 127
 QY 589 CTGAATCCAAAGATGCTG-----CTTCCCTCAGATTTGTT----- 624
 DB 128 MethrasnlysmelleuhslysrphelaleuProlethyrllelehslyrlelph 147
 QY 625 -----TTCTTCTTACGATCCATTAATCCCTCTGGAGACAGACAGTACTAG 678
 DB 148 ProthrilevalleuProserleuvallylsleuPro----- 159
 QY 679 AACTCTCAATTTCTTTGTTTGTGTTTAAAGACAGACTCTGTCAAAAAAGGACA 738
 DB 160 -----AspGlnGlnThrGlylyslalasn 167
 QY 739 TTTATCATTAATCATCTTATAGAGCCCAATTTCTTATCGAAGCACTGTTTTTTT 798
 DB 168 PheleuLeu----- 170
 QY 799 TTTTAAAGCTTAAGTACTGATGTCACAGACAAATATTTCTGATCAGATAGCCCTGT 858
 DB 171 -----LystyrlyGlyCys-----ValProProtyr 178
 QY 859 CAACAGTAGCAAAATGTGCTTTCATTAAGTGGGAAGAAAGACAGATTTTAAAGTAAT 915
 DB 179 Valaspleucluarlyvalphetyrleuilellethrlsarsglyrphelulethrcys 198
 QY 916 -----TTTTGG-----GAGACTGATTTG 933
 DB 199 AlavalphellecysThrmetPheAlaGluvalTrrPhePheAlaleuvalThrAspArg 218
 QY 934 AGTAATAATTAAGCTGTCTGCTCCCTTAAGAAAAAAACCTTCCACCTTAC----- 987
 DB 219 leuileuylsglnmetThrlysthrmetSerGlnlysthrphespleuhslyls 238
 QY 988 -----TGTCATTTTATATCCCTTATGTTCCAAAGTTAATTAATTTATTTCTGAT 1038
 DB 239 PheGlnArgAlaPhelelleuGlnleuLeuPro-----PhelellevalPheleuPro 257
 QY 1039 ATGCTTTTATACCAAGAGCCCTTATCAGGAGCTTCAGACAGACACTATAGCAGAC 1098
 DB 258 lleserlyrileglyvalthrlyscysleseerlytyrhisnGlnPhePhe----- 274
 QY 1099 CATACCAACCAATACCAACATATATGTTTATATATCATCATAGTACGACATGA 1158
 DB 275 -----AsnAsnleuThrleuilelleleleSerSerHisGly----- 286
 QY 1159 TTATTGAGGCTTGATGGCAATATAGCACTTTCATATCTTCATCTTTCATATCCAT 1218
 DB 287 PhePheSerThrlellelelleuilellelelelelelelelelelelelelelele 306
 QY 1219 ATCAGACTACTACACTTTTGTGNAAG---TCATCTAGACAGAAATCGAATGAAGACC 1275
 DB 307 lle-----PheProCysleuylsargPheSerSerleAlathrSerSerThrAlaPro 324
 QY 1276 TATATTTTCTGATGATCT 1293
 DB 325 GlnSerleuemetillethr 330

Search completed: June 24, 2003, 19:24:00
 Job time : 112.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2003, 03:14:56 ; Search time 3443 Seconds

(Without alignments)
11250.605 Million cell updates/sec

Title: US-09-830-244B-2

Perfect score: 1331

Sequence: 1 ccatgttttagaatcaaa.....gcctttccaatgctactg 1331

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Geneml:*

1: gb_ba:*

2: gb_hlg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vl:*

30: em_hlg_hum:*

31: em_hlg_inv:*

32: em_hlg_other:*

33: em_hlg_mus:*

34: em_hlg_pln:*

35: em_hlg_rnd:*

36: em_hlg_mam:*

37: em_hlg_vrt:*

38: em_sy:*

39: em_hlgo_hum:*

40: em_hlgo_mus:*

41: em_hlgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	704	52.9	133984	9	AL512306	AL512306 Human DNA
C 2	396	29.8	153023	2	AC021462	AC021462 Homo sapi
C 3	361.2	27.1	153023	2	AC021462	AC021462 Homo sapi
C 4	349.2	26.2	146805	2	AL450424	AL450424 Homo sapi
C 5	253	19.0	542	9	HUMSTRNA	M18371 Human stath
C 6	253	19.0	552	9	HUMSTRNA	M18078 Human stath
C 7	253	19.0	1584	9	AK092678	AK092678 Homo sapi
C 8	176.2	13.2	297	11	G05711	G05711 human STS
C 9	145.4	10.9	1498	11	G26643	G26643 human STS
C 10	145.4	10.9	4723	9	HUMSTRNA2	M32639 Human saliv
C 11	145.4	10.9	141568	9	AC063956	AC063956 Homo sapi
C 12	145.4	10.9	161549	2	AC024676	AC024676 Homo sapi
C 13	86.2	6.5	524	9	HUMHIS2X	M26665 Human hist
C 14	86.2	6.5	558	9	BC009791	BC009791 Homo sapi
C 15	81	6.1	480	9	HUMHIS1X	M26664 Human hist
C 16	81	6.1	566	9	BC017835	BC017835 Homo sapi
C 17	78.8	5.9	491	9	HUMHHRPA	M18372 Human hist
C 18	77.2	5.8	209410	2	AL805933	AL805933 Mus muscu
C 19	74.6	5.6	9881	6	AX281311	AX281311 Sequence
C 20	72	5.4	193445	2	AL845502	AL845502 Mus muscu
C 21	64.8	4.9	81962	2	AC128396	AC128396 Rattus no
C 22	64.8	4.9	246173	2	AC094026	AC094026 Rattus no
C 23	62.2	4.7	183969	2	AC069037	AC069037 Homo sapi
C 24	59	4.4	1764	9	HUMHIS103	L05512 Human hist
C 25	59	4.4	185969	2	AC069037	AC069037 Homo sapi
C 26	55.8	4.2	9881	6	AX281312	AX281312 Sequence
C 27	52.4	3.9	164399	3	PFMAL3P6	PFMAL3P6 Plasmid
C 28	52.2	3.9	128635	2	AC108074	AC108074 Homo sapi
C 29	52.2	3.9	154472	2	AC109457	AC109457 Homo sapi
C 30	52.2	3.9	207558	2	AC114778	AC114778 Homo sapi
C 31	51.6	3.9	170141	2	AC092022	AC092022 Homo sapi
C 32	51.6	3.9	171044	2	AC092787	AC092787 Homo sapi
C 33	51.6	3.9	194874	2	AC080090	AC080090 Homo sapi
C 34	51.6	3.9	274349	2	AC093623	AC093623 Homo sapi
C 35	51.2	3.8	12029	3	AE001373	AE001373 Plasmid
C 36	51	3.8	182362	9	AL160033	AL160033 Human DNA
C 37	48.6	3.7	1141	6	AX083744	AX083744 Sequence
C 38	48.6	3.7	157051	9	AC011095	AC011095 Homo sapi
C 39	48.6	3.7	167830	2	AC027768	AC027768 Homo sapi
C 40	48.6	3.7	173765	9	AC090457	AC090457 Homo sapi
C 41	48.4	3.6	6056	6	AX345928	AX345928 Sequence
C 42	48.4	3.6	121709	9	AC093783	AC093783 Homo sapi
C 43	48.4	3.6	12989	9	AC005681	AC005681 Homo sapi
C 44	48.4	3.6	174277	9	AC093826	AC093826 Homo sapi
C 45	48	3.6	61215	2	AC090357	AC090357 Homo sapi

ALIGNMENTS

RESULT 1

AL512306

LOCUS Human DNA sequence from clone Rpl1-430C7 on chromosome 1, complete

DEFINITION

ACCESSION AL512306

VERSION AL512306

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 133984)

AUTHORS Bray-Allen,S.

TITLE Direct Submission

|||||
Db 4265 GGCCTAATTTGGGAACAGTTTCTGATGCTTTGAGAACTACTTCTTTGACAGAAA 4206
QY 301 TTTCATTCCTGCTGCAATGCTATTCCTTATCCCTTTATAGACCATTTGATTTCTTCC 360
|||||
Db 4205 TTTCATTCCTGCTGCAATGCTATTCCTTATAGACCATTTGATTTCTTCC 4146
QY 361 TTTTGTGGGAATGTCCTATTCATTTGATCT 396
|||||
Db 4145 TTTTGTGGGAATGTCCTATTCATTTGATCT 4110
|||||
RESULT 3
AC021462
LOCUS
DEFINITION
AC021462 153023 bp DNA linear HTG 03-APR-2000
Homo sapiens clone RP11-276C1, WORKING DRAFT SEQUENCE, 21 unordered
pieces.
AC021462 GI:7387343
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
Homo sapiens.
ORGANISM
Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 153023)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barne, N., Becker, R., Beld, F.,
Boguslavsky, L., Bouckhagalter, B., Brown, A., Burkett, G., Castle, A.,
Choe, Y., Collangelo, M., Collins, S., Collamore, A., Cooke, P.,
Dearrellano, K., Dewar, K., Donlin, M., Doyle, M., Feneator, J.,
Ferris, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,
Gardner, S., Grant, G., Hagos, B., Harford, A., Horton, L.,
Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Lander, T., Lehoczeky, J., Levine, R., Lieu, C., Liu, G., Locke, K.,
Macdonald, P., Margulis, N., McEwan, P., McGurk, A., McKernan, K.,
McPherson, R., Meldrum, J., Menus, L., Morrow, J., Naylor, J.,
Norman, C.H., O'Connor, T., O'Donnell, P., Oliver, J.M., Peterson, K.,
Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomson, N.,
Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W., J.,
Zimmer, A. and Zody, M.
TITLE
JOURNAL
COMMENT
Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 3, 2000 this sequence version replaced g1:7230200.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 15201
Center clone name: 276_C1
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; Version 0.960731
Consensus quality: 141815 bases at least Q40
Consensus quality: 147368 bases at least Q20
Consensus quality: 149481 bases at least Q20
Insert size: 157000; agarose-fp
Insert size: 151023; sum-of-ctrls
Quality coverage: 4.3 in Q20 bases; agarose-fp
Quality coverage: 4.5 in Q20 bases; sum-of-ctrls

NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1015: contig of 1015 bp in length
1016 1115: gap of 100 bp
1116 2276: contig of 1161 bp in length
2277 2376: gap of 100 bp
2377 4009: contig of 1633 bp in length
4010 4109: gap of 100 bp
4110 5737: contig of 1628 bp in length
5738 5837: gap of 100 bp
5838 6939: contig of 1102 bp in length
6940 7039: gap of 100 bp
7040 9504: contig of 2465 bp in length
9505 9604: gap of 100 bp
9605 11734: contig of 2130 bp in length
11735 11834: gap of 100 bp
11835 14111: contig of 2277 bp in length
14112 14211: gap of 100 bp
14212 16973: contig of 2762 bp in length
16974 17073: gap of 100 bp
17074 19588: contig of 2515 bp in length
19589 19688: gap of 100 bp
19689 22275: contig of 2587 bp in length
22276 22375: gap of 100 bp
22376 25823: contig of 3448 bp in length
25824 25923: gap of 100 bp
25924 31307: contig of 5384 bp in length
31308 31407: gap of 100 bp
31408 34863: contig of 3456 bp in length
34864 34963: gap of 100 bp
34964 41382: contig of 6419 bp in length
41383 41482: gap of 100 bp
41483 49024: contig of 7542 bp in length
49025 49124: gap of 100 bp
49125 58672: contig of 9548 bp in length
58673 58772: gap of 100 bp
58773 69622: contig of 10850 bp in length
69623 69722: gap of 100 bp
69723 88191: contig of 18469 bp in length
88192 88291: gap of 100 bp
88292 107084: contig of 18793 bp in length
107085 107184: gap of 100 bp
107185 153023: contig of 45839 bp in length.
FEATURES
SOURCE
1. 153023
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-276C1"
/clone_id="RPC1-11 Human Male BAC"
1. 1015
/note="assembly-fragment"
1116. 2276
/note="assembly-fragment"
2377. 4009
/note="assembly-fragment"
4110. 5737
/note="assembly-fragment"
5838. 6939
/note="assembly-fragment"
7040. 9504
/note="assembly-fragment"
9605. 11734
/note="assembly-fragment"
11835. 14111
/note="assembly-fragment"
14212. 16973
/note="assembly-fragment"

BASE COUNT	41494	a	35192	c	34573	g	39756	t	2008	others
ORIGIN	.									

QY	709	GACAGAGACTCTGTGCCAAAAAAGGACATTTATCATTTATAACATCTTATTAGAGCCCC	768
Db	280	GACAGAGACTCTGTCTCCAAAAAAGGACATTTATCATTTATAACATCTTATTAGAGCCCC	339
QY	769	TAATTTCTTATCTGAAGGCACGTGTTTTTTTTTTTAAACAGTTAAAGTACGATGTAACAG	828
Db	340	TAATTTCTTATCTGAAGGCACGTGTTTTTTTTTTTAAACAGTTAAAGTACGATGTAACAG	399
QY	829	ACAAATATTTCTGATCAGATAGTCCCTGTGCAACAGTAGCAAAATGTGTTTCATTAAGTG	888
Db	400	ACAAATATTTCTGATCAGATAGTCCCTGTGCAACAGTAGCAAAATGTGTTTCATTAAGTG	459
QY	889	GGAGAAAAACGACATTTTAACTTTTGGGAGACTGATTTGATATATATAAACT	948
Db	460	GGAGAAAAACGACATTTTAACTTTTGGGAGACTGATTTGATATATATAAACT	519
QY	949	CTGTCTCCCTTAAAGAAAAAACCCTCCACCTTACTGTGTCATTTATATCCCTTA	1008
Db	520	CTGTCTCCCTTAAAGAAAAAACCCTCCACCTTACTGTGTCATTTATATCCCTTA	579
QY	1009	GTTCCAAAGTTAAATATCTTATTTCTGATATATGCTTTATACCAAGAGCCTTATTCAGC	1068
Db	580	GTTCCAAAGTTAAATATCTTATTTCTGATATATGCTTTATACCAAGAGCCTTATTCAGC	639
QY	1069	CAGTTCGAGAACCAACCATATA	1090
Db	640	CCTGTGTACTACAGTATCTTA	661

RESULT 4	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
AL450424			AL450424				
		146805 bp DNA linear HTG 12-SEP-2001					
		Homo sapiens chromosome 1 clone RP11-563116, *** SEQUENCING IN PROGRESS ***, 25 unordered pieces.					
			AL450424				
			AL450424.3	GI:12331147			
			HTG; HTGS_PHASE1; HTGS_CANCELLED.				
			human.				
			Homo sapiens				
			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

Mammalia: Eutheria: Primates: Catarrhini, Hominoidea: Homo.
1 (bases 1 to 146805)
Mclay, K
Direct Submission
Submitted (11-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jan 22, 2001 this sequence version replaced gl:11493361.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: ba563116
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L081752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 13533 bases at least Q40
Consensus quality: 13636 bases at least Q30
Consensus quality: 141952 bases at least Q20
Insert size: 144405; sum-of-connigs
Insert size: 171032; 2.3% error; agarose-fp
Quality coverage: 3.60x in Q20 bases; sum-of-connigs Quality
coverage: 3.25x in Q20 bases; agarose-fp
-----
** NOTE: This is a 'working draft' sequence. It currently
** consists of 25 connigs. The true order of the pieces
** is not known and their order in this sequence record is
** arbitrary. Gaps between the connigs are represented as
** runs of N, but the exact sizes of the gaps are unknown.
** This record will be updated with the finished sequence
** as soon as it is available and the accession number will
** be preserved.
1
5105: connig of 5105 bp in length
5106 5205: gap of 100 bp
5206 10570: connig of 5365 bp in length
10571 10670: gap of 100 bp
10671 17626: connig of 6956 bp in length
17627 17726: gap of 100 bp
17727 20478: connig of 2752 bp in length
20479 20578: gap of 100 bp
20579 23035: connig of 2457 bp in length
23036 23135: gap of 100 bp
23136 25308: connig of 2173 bp in length
25309 25408: gap of 100 bp
25409 30469: connig of 5061 bp in length
30470 30569: gap of 100 bp
30570 32688: connig of 2119 bp in length
32689 32788: gap of 100 bp
32789 36187: connig of 3399 bp in length
36188 36287: gap of 100 bp
36288 48988: connig of 12701 bp in length
48989 49088: gap of 100 bp
49089 67333: connig of 18245 bp in length
67334 67433: gap of 100 bp
67434 82317: connig of 14884 bp in length
82318 82417: gap of 100 bp
82418 84614: connig of 2197 bp in length
84615 84714: gap of 100 bp
84715 90326: connig of 5612 bp in length
90327 90426: gap of 100 bp
90427 92647: connig of 2221 bp in length
92648 92748: gap of 100 bp
92748 97690: connig of 4943 bp in length
97691 97790: gap of 100 bp
97791 100799: connig of 3009 bp in length
100800 100899: gap of 100 bp
100900 104202: connig of 3303 bp in length
104203 104303: gap of 100 bp
104303 107157: connig of 2855 bp in length
107158 107257: gap of 100 bp

```

```

*
* 107258 110300: contig of 3043 bp in length
* 110301 110400: gap of 100 bp
* 110401 113955: contig of 3555 bp in length
* 113956 114055: gap of 100 bp
* 114056 118801: contig of 4746 bp in length
* 118802 118901: gap of 100 bp
* 118902 129831: contig of 10930 bp in length
* 129832 137545: contig of 7614 bp in length
* 137546 137645: gap of 100 bp
* 137646 146805: contig of 9160 bp in length.
FEATURES
source
1. 146805
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone_1lb="RP11-563116"
/clone_1lb="RP11-11.2"
1. 5105
/note="assembly_fragment:00368
clone_end:T7
vector_side:left"
5206. 10570
/note="assembly_fragment:00594
fragment_chain:1"
10671. 17626
/note="assembly_fragment:00786
fragment_chain:1"
17727. 20478
/note="assembly_fragment:00844
fragment_chain:1"
20579. 23035
/note="assembly_fragment:00011
fragment_chain:1"
23136. 25308
/note="assembly_fragment:01029
fragment_chain:1"
25409. 30469
/note="assembly_fragment:00148
fragment_chain:2"
30570. 32688
/note="assembly_fragment:00961
fragment_chain:2"
32789. 36187
/note="assembly_fragment:00310
fragment_chain:3"
36288. 48988
/note="assembly_fragment:01150
fragment_chain:3"
49089. 67333
/note="assembly_fragment:00444
fragment_chain:4"
67434. 82317
/note="assembly_fragment:00140
fragment_chain:4"
82418. 84614
/note="assembly_fragment:00657
fragment_chain:5"
84715. 90326
/note="assembly_fragment:00757
fragment_chain:5"
90427. 92647
/note="assembly_fragment:00677
fragment_chain:6"
92748. 97690
/note="assembly_fragment:00248
fragment_chain:6"
97781. 100799
/note="assembly_fragment:00091"
100900. 104202
/note="assembly_fragment:00538"
104303. 107157
/note="assembly_fragment:00675"

```

```

misc_feature 107258..110300
/note="assembly_fragment:00695"
misc_feature 110401..113955
/note="assembly_fragment:01316"
misc_feature 114056..118801
/note="assembly_fragment:01292
fragment_chain:7"
misc_feature 118902..129831
/note="assembly_fragment:00551
fragment_chain:7"
misc_feature 129832..137545
/note="assembly_fragment:01410
fragment_chain:7"
misc_feature 137646..146805
/note="assembly_fragment:00763
fragment_chain:7
clone_end:SP6
vector_side:right"
BASE COUNT 37508 a 33240 c 34076 g 39557 t 2424 others
ORIGIN

```

```

Query Match 26.2% Score 349.2; DB 2; Length 146805;
Best Local Similarity 96.3%; Pred. No. 7.9e-65;
Matches 368; Conservative 0; Mismatches 13; Indels 1; Gaps 1;

QY 709 GACAGAGACTGTCTCTCAAAAAAAGACATTATCATTAATATTAAGAGCCCC 768
DB 84740 GACAGAGACTGTCTCTCAAAAAAAGACATTATCATTAATATTAAGAGCCCC 84799
QY 769 TAATTCCTATCTGAAGCAGCTGTTTTTTTAAACAGTTAACTGATGATGACAG 828
DB 84800 TAATTCCTATCTGAAGCAGCTGTTTTTTTAAACAGTTAACTGATGATGACAG 84859
QY 829 ACAATATTTCTGATCAGATAGTCCCTGTCAACAGTACAAATGCTTCATTAAGTG 888
DB 84860 ACAATATTTCTGATCAGATAGTCCCTGTCAACAGTACAAATGCTTCATTAAGTG 84919
QY 889 GGAAGAAAAACGACATTTTAAAGTAATCTTTGGGAGACTGATTGAGTAATATAAACT 948
DB 84920 GGAAGAAAAACGACATTTTAAAGTAATCTTTGGGAGACTGATTGAGTAATATAAACT 84978
QY 949 CTGGTCTCCCTTAAGAAAAAACCCTTCCACCTTCTGTCATTTATATCCCTTA 1008
DB 84979 CTGGTCTCCCTTAAGAAAAAACCCTTCCACCTTCTGTCATTTATATCCCTTA 85038
QY 1009 GTTCCAAGTTAATTATCTTATTTCTGATATTGCTTTATACCAAGAGCTTATCAGC 1068
DB 85039 GTTCCAAGTTAATTATCTTATTTCTGATATTGCTTTATACCAAGAGCTTATCAGC 85098
QY 1069 CAGTTCAGAACACCACTATA 1090
DB 85099 CAGTTCAGAACCACTATCTTA 85120

```

```

RESULT 5
HMMSTRNA 542 bp mRNA linear PRI 13-JAN-1995
LOCUS Human statherin mRNA, complete cds.
DEFINITION Human statherin mRNA, complete cds.
ACCESSION M18371
VERSION M18371.1 GI:338610
KEYWORDS statherin.
SOURCE Human female submandibular gland, cDNA to mRNA, clone pBRHSP98.2.
ORANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 542)
AUTHORS Dickinson,D.P., Ridall,A.L. and Levine,M.J.
Human submandibular gland statherin and basic histidine-rich
peptide are encoded by highly abundant mRNA's derived from a common
ancestral sequence
JOURNAL Biochem. Biophys. Res. Commun. 149 (2), 784-790 (1987)
MEDLINE 88106506
PUBMED 3426601

```


Magatsuma, M., Murakawa, K., Kanehori, K., Sugiyama, A., Kawakami, B.,
Suzuki, Y., Sugano, S., Nagahara, K., Masuno, Y., Nagai, K. and
Isogai, T.
NEO human cDNA sequencing project
Unpublished
2 (bases 1 to 1584)
Isogai, T. and Yamamoto, J.
Direct Submission
Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'-3' and one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.
Location/Qualifiers

COMMENT

FEATURES

Source
1. 1584
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="SALGI100107"
/tissue_type="salivary gland"
/clone_lib="SALGI1"
/note="cloning vector: pME18SFL3"
89. 997
/note="unnamed protein product"
/codon_start=1
/protein_id="BAC03943.1"
/db_xref="GI:21751328"
/translation="MGDSEMLKLPVQKCEHKLWKARLSGYEALKIFQKIDKPS
EMSKFPLIKKFWTDSNAVVLKGLAEALVYVNAHVAGKCPVSVSVSGVFNQPK
AKKEIGIEICLMYIEIEKKEGAVOEELIKLGNKPKIVATIEFLKRALSGSKII
LKPIIKVLKLFESREKARVDEAKILANETIRWIDRLPQLQININVOLEEM
VKPTAPRPTRFLRSQLELKLKLEQDSAGDAGGDDGDEVPQIDAYELLEAVEI
LSKLPTEMTKLKNGKREKRPWSL"

CDS

BASE COUNT 526 a 298 c 331 g 429 t
ORIGIN

Query Match 19.0%; Score 253; DB 9; Length 1584;
Best Local Similarity 88.4%; Pred. No. 6.6e-44;
Matches 274; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

1022 TTATCTTATTTCTGATATTCCTTTATACCAAGAGCCTTTCAGCCAGTTCCAGACA 1081
1174 TTGCGTAGAATTGGAAGATTGGTTATGGTATGCGCCTTATCAGCCAGTTCCAGACA 1233
1082 ACCACTATTCGACCAACCAATACCAACCAATACCAATATACGTTTAAATATCATCA 1141
1234 ACCACTATTCGACCAACCAATACCAACCAATACCAATATACGTTTAAATATCATCA 1293
1142 GTAACTGACGACATGATATGAGGCTTGATGAGCAATACGACTTTACATCCATATT 1201
1294 GTAACTGACGACATGATATGAGGCTTGATGAGCAATACGACTTTACATCCATATT 1353
1202 CTCATCTTTCATATCATATCACACTACTACCACTTTTGTNAGATCATCTAAGACATG 1261
1354 CTCATCTTTCATATCATATCACACTACTACCACTTTTGTNAGATCATCTAAGACATG 1413
1262 GGAATGTAAACCCATATATTTACGCACTCTTTGGTTCCAGATCTGCTTTTCCA 1321
1414 CAATAGTAAACACATATATTTACGCACTCTTTGGTTCCAGATCTGCTTTTCCA 1473
1322 ATGTCACTTG 1331
1474 TTGTCACTTG 1483

RESULT 8
G05711
LOCUS G05711 297 bp DNA linear STS 19-OCT-1995

DEFINITION human STS WI-7844, sequence tagged site.
ACCESSION G05711
VERSION G05711.1 GI:858956
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE Homo sapiens STS derived from sequences in dbEST and the UniGene collection.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 297)
Hudson, T.
Whitehead Institute/MIT Center for Genome Research; Physically
Mapped ESTs
Unpublished (1995)

JOURNAL

Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
Whitehead Institute for Biomedical Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu

Primer A: ACTGACGACATGATTTATGACG
Primer B: AATAATGCTTGAATTTATTC
STS size: 286
PCR Profile:

Presoak:
Denaturation: 56 degrees C
Annealing: 56 degrees C
Polymerization:
PCR Cycles: 35
Thermal Cycler:

Protocol:
Template: 10 ng
Primer: each 5 pm
dNTPs: each 4 mM
Tag Polymerase: 0.025 units/ul
Total Vol: 20 ul

Buffer:
MgCl2: 1.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 9.3

Prepared with primer pairs derived from M18371 -- Unigene.

FEATURES
source
1. 297
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/map_F_3: 793_E_2: 794_A_9: 886_B_3: 960_D_2:
972_E_4: 750_D_11: 855_E_12: 882_G_11: 931_G_10: 395.8 CR
from top of Chr4 linkage group"

STS
primer_bind 12. 297
primer_bind 12. 34
BASE COUNT 99 a 56 c 37 g 105 t
ORIGIN

Query Match 13.2%; Score 176.2; DB 11; Length 297;
Best Local Similarity 92.9%; Pred. No. 2.2e-27;
Matches 184; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

1134 TATCATGATTAAGTACGACATGATTTATGAGGCTTGATGCAATACGACTTTTACA 1193
1 TATCATGATTAAGTACGACATGATTTATGAGGCTTGATGCAATACGACTTTTACA 60
1194 TCCATATTCATCTCTTTATACCATATACGACTACTTCTTTTGTNAGATCATGTAA 1253
61 TCCATATTCATCTCTTTATACCATATACGACTACTTCTTTTGTNAGATCATGTAA 120
1254 GAGCAATGCAATGTAAACCCATATTTACTGATACCTTTGGTTCCAGATCTTGC 1313


```

Db      121 GAGCAATGCAATGAAAAACACTATATTTACTGTATACCTTTGTTTCAGAGATACCTGC 180
QY      1314 CTTTTCAGTGTCACTTG 1331
Db      181 CTTTTCAGTGTCACTTG 198

RESULT 9
LOCUS   G26643
DEFINITION human STS STS_M32639, sequence tagged site.
ACCESSION G26643
VERSION  G26643.1 GI:1348875
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE   Homo sapiens STS derived from sequences in dbEST and the UniGene collection.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS  1 (bases 1 to 1498)
TITLE    Whitehead Institute/MIT Center for Genome Research; Physically Mapped STSS
JOURNAL  Unpublished (1995)
COMMENT  Contact: Thomas Hudson
          Whitehead Institute/MIT Center for Genome Research
          Whitehead Institute for Biomedical Research
          9 Cambridge Center, Cambridge MA 02142 USA
          Tel: 617 252 1900
          Fax: 617 252 1902
          Email: thudson@genome.wi.mit.edu

Primer A: GCACATGATTATTGAGTAAAGATGG
Primer B: ACTGCCTTCAACCTACACAGG
STS size: 210
PCR Profile:
  Presoak:
  Denaturation:
  Annealing: 56 degrees C
  Polymerization:
  PCR Cycles: 35
  Thermal Cycler:
  Protocol:
    Template: 10 ng
    Primer: each 5 pm
    dNTPs: each 4 mM
    Tag Polymerase: 0.025 units/ul
    Total Vol: 20 ul

Buffer:
  MgCl2: 1.5 mM
  KCl: 50 mM
  Tris-HCL: 10 mM
  pH: 9.3

FEATURES
  source
    Derived from dbEST (genbank accession M32639).
    Location/Qualifiers
      1..1498
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /map="473.8 cr from top of Chr4 linkage group"
  STS
    Primer_bind
      21..45
    Primer_bind
      complement(209..230)
  BASE COUNT  511 a 229 c 244 g 514 t
  ORIGIN

Query Match      10.9%; Score 145.4; DB 11; Length 1498;
Best Local Similarity 90.1%; Pred. No. 6,1e-21;
Matches 155; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY      1160 TATTGAGCTTGATTTGGCAATACGACTTCTACATTCATATTCATCTTTCATATACCATTA 1219
Db      1234 TGTGACAGGCTTGATTTGGCAATACGACTTCTACATTCATATTCATCTTTCATATACCATTA 1293
QY      1220 TCACACTACACACCTTTTGTNAGATCATCTAAGACCAATGCAATGTAAACCCATATA 1279
Db      1294 TCACACTACACACCTTTTGTNAGATCATCTAAGACCAATGCAATGTAAACCCATATA 1353
QY      1280 ATTACTGATACCTTTGGTCCAGATACCTTGCCTTTTCATATTCACCTTG 1331
Db      1354 ATTACTGATACCTTTGGTCCAGATACCTTGCCTTTTCATATTCACCTTG 1405

RESULT 10
LOCUS   HUMSTAT2
DEFINITION Human salivary statherin gene, exons 2-6.
ACCESSION M32639
VERSION  M32639.1 GI:338504
KEYWORDS statherin.
SEGMENT 2 of 2
SOURCE   Human (individuals #563, #8136, and J.F.) fibroblast, cell line #563, DNA, clones 1-3.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS  1 (bases 1 to 4723)
TITLE    Sabatini, L.M., He, Y.Z. and Azen, E.A.
          structure and sequence determination of the gene encoding human salivary statherin
JOURNAL  Gene 89 (2), 245-251 (1990)
MEDLINE  90323623
PUBMED  2373369
COMMENT  Draft entry and computer-readable sequence for [1] kindly submitted by L.M.Sabatini, 03-JAN-1990.
FEATURES
  source
    Location/Qualifiers
      1..4723
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /map="4q11-q13"
        /join(M31077.1:1601..1931,1..352)
        /gene="STATH"
        /note="STATH mRNA and introns"
      <1..4309
        /note="STATH"
      <1..286
        /gene="STATH"
        /note="STATH intron A"
        /join(302..352,1532..1552,1645..1674,2728..2814)
        /note="statherin precursor"
        /codon_start=1
        /protein_id="AA060593.1"
        /db_xref="GI:338506"
        /translation="MKELVFAFILALVMYMGADSSSEKFLRRIGRGYGYGPYPVP
          EQPLYPQYPOYPOYPOYTP"
        /join(302..352,1532..1537)
        /note="statherin signal peptide"
        /join(1538..1552,1645..1674,2728..2811)
        /product="statherin"
      <302..352
        /gene="STATH"
        /note="statherin precursor, (first translated exon);
          G00-120-391"
        /number=2
        353..1531
        /note="STATH intron B"
        1532..1552
        /number=3
        1553..1644
        /note="STATH intron C"
        1645..1674
        /number=4
        1675..2727
        /note="STATH intron D"

```

```

exon          2728..>2814
              /note="stathlerin precursor"
Intron        /number=5
              2848..4051
              /note="STATH Intron E"
polyA_signal  4294..4299
BASE COUNT    1570 a 797 c 722 g 1634 t
ORIGIN        About 1.8 kb after segment 1.

Query Match   10.9% Score 145.4; DB 9; Length 4723;
Best Local Similarity 90.1% Pred. NO. 4.8e-21;
Matches 155; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 1160 TATTGAGCTTGATGGCAATATGACATTCATATTCATCTTCATACACATA 1219
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4045 TGTGACAGCTTGATGGCAATATGACATTCATATTCATCTTCATACACATA 4104

QY 1220 TCACACTACTACACTCTTTTGTNAGATCATCTAGAGCAATGCGAATGTAACCCATTA 1279
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4105 TCACACTACTACACTCTTTTGTNAGATCATCTAGAGCAATGCGAATGTAACCCATTA 4164

QY 1280 ATTACTAGATGATCTTGTGTCACAGATCTTGCCTTTTCATATGTCATCTT 1331
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4165 ATTACTAGATGATCTTGTGTCACAGATCTTGCCTTTTCATATGTCATCTT 4216

RESULT 11
AC063956      141568 bp      DNA      linear      PRI 25-AUG-2000
LOCUS        Homo sapiens 4 BAC RP11-529K3 (Roswell Park Cancer Institute Human
AC063956      BAC library) complete sequence.
VERSION      AC063956.7 GI:9910030
KEYWORDS     HTG.
SOURCE       Homo sapiens.
ORGANISM     Homo sapiens.
REFERENCE    Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
              1 (bases 1 to 141568)
              Alabrooks,S.L., Amaralunga,H.C., Are,J.R., Banks,T., Barberia,J.,
              Benton,J., Blinige,K., Blankenburg,K., Bonnin,D., Bouck,J.,
              Bowler,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
              Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
              Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
              Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
              Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
              Day-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
              Dem,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
              Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
              Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
              Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
              Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
              Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
              He,X., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M.,
              Hollway,C., Hollins,B., Homsl,F., Howard,S., Huber,J., Hulyk,S.,
              Hume,J., Ioshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y.,
              Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U.,
              Kling,L., Korvah,J., Kovar,C., Krawczyk,J., Kureshi,A., Landry,N.,
              Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lien,C.,
              Liu,J., Liu,W., Louisedge,H., Lozado,R.J., Lu,X., Lucier,A.,
              Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R.,
              Martindale,A., Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P.,
              Meador,M., Mel,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T.,
              Mohabbat,K., Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M.,
              Neal,D., Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
              Nguyen,N., Nickerson,E., Nnokenkwo,S., Ogun,M., Okunolu,G.,
              Orangunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
              Peters,L., Pickens,R., Plimus,E., Pu,L.L., Quiles,M., Ren,Y.,
              Rives,M., Rojas,A., Rojuboan,I., Rolfe,M., Ruiz,S., Savery,G.,
              Scherer,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I.,
              Sodergren,E., Sonalike,T., Sparks,A., Stanley,H., Stone,H.,
              Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H.,

```

```

TITLE        Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
              Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wall,R.,
              Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wellington,S.,
              Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K.,
              Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Kucherlapati,R.,
              Naylor,S.L. and Gibbs,R.
              Naylor,S.L. and Gibbs,R.
DIRECT SUBMISSION
Unpublished
2 (bases 1 to 141568)
REFERENCE    Morley,K.C.
AUTHORS      Morley,K.C.
TITLE        Direct Submission
JOURNAL      Submitted (22-APR-2000) Human Genome Sequencing Center, Department
              of Molecular and Human Genetics, Baylor College of Medicine, One
              Baylor Plaza, Houston, TX 77030, USA
              3 (bases 1 to 141568)
              Morley,K.C.
              Direct Submission
              Submitted (25-AUG-2000) Human Genome Sequencing Center, Department
              of Molecular and Human Genetics, Baylor College of Medicine, One
              Baylor Plaza, Houston, TX 77030, USA
              On Aug 25, 2000 this sequence version replaced g1:9795448.
              INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
              gc-help@bcm.tmc.edu

```

COMMENT

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: http://gc.bcm.tmc.edu:8088/quality_intro/genbank_annotation..html.

QUALSTAT-REPORT

```

----- Summary Statistics -----
Contig Length: 141568
Phrap values in estimate: 140895
Average error rate (BCM-Phrap estimate): 1.4137e-05
Fraction of Phrap values less than 40 : 0.00715426
Number of consensus changing edits: 10
Number of N's in consensus : 0

```

```

----- Consensus changing edits -----
Position Original+Context Edited+Context
23165 cctgaattct(n)tcctgtgcaa cctgaattct(t)tcctgtgcaa

```


Db 126696 ATTACTGTATCTTTCTTTCAGAGTACTGCTTTTCATTTCTCATTG 126747

RESULT 13
LOCUS HUMHIS2X 524 bp mRNA linear PRI 31-DEC-1994
DEFINITION Human histatin 2 (HIS2) mRNA, complete cds.
ACCESSION M26665
VERSION M26665.1 GI:292145
KEYWORDS histatin 2
SOURCE Homo sapiens parotid gland cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE Sabatini, L.M. and Azen, E.A.
AUTHORS 1 (bases 1 to 524)
TITLE Histatins, a family of salivary histidine-rich proteins, are
JOURNAL encoded by at least two loci (HIS1 and HIS2)
MEDLINE Biochem. Biophys. Res. Commun. 160 (2), 495-502 (1989)
PUBMED 89246491
2719677

FEATURES
source location/Qualifiers
1..524 /organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="parotid gland"
1..524 /gene="HIS2"
72..227 /gene="HIS2"
/codon_start=1
/product="histatin 2"
/protein_id="AA58646.1"
/db_xref="GI:292146"
/translation="MKFFVALLALMLSMTGADSHAKRHGKRRFHEKHSHRGYR
SNLYDN"
506..511
polyA_signal /gene="HIS2"
BASE COUNT 157 a 100 c 87 g 180 t
ORIGIN

Query Match 6.5%; Score 86.2; DB 9; Length 524;
Best Local Similarity 62.5%; Pred. No. 3.2e-08;
Matches 150; Conservative 0; Mismatches 89; Indels 1; Gaps 1;

QY 1093 CACACCATACCAACCAATACCAATATGTTTAATATCATCTAGTACGAG 1152
Db 187 CACATCGAGGCTATGATCAATATCTGTATGACATGATTCCTTCATATCATGGG 246
QY 1153 ACATGATTTATGAGGCTTGATGGCAATATGACATTCATATTCATCTTCA 1212
Db 247 GCATGATTTATGAGGCTTGATGGCAATATGACATTCATATTCATCTTCA 306
QY 1213 TACCATATCAGACTACTACCATCTTTGTGATGATCTCTTAAGCAATGC-GAATGTAA 1271
Db 307 TACCGGATCAGACTACTACCATCTTTGTGATGATCTCTTAAGCAATGC-AATGAAG 366
QY 1272 ACCCTAATATTTACTGATATCTTTGTTCCAGATATCTGCTTTCATGTCACCTG 1331
Db 367 AATATACATGATTTAGTGAATCTGTGTTTCAGATATCTCCCTTCCTAATATTCATTGG 426

RESULT 14
LOCUS BC009791 558 bp mRNA linear PRI 12-JUL-2001
DEFINITION Homo sapiens, histatin 3, clone MGC:13578 IMAGE:4293405, mRNA,
complete cds.
ACCESSION BC009791
VERSION BC009791.1 GI:14602560
KEYWORDS MGC.
SOURCE MGC. sapiens.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE 1 (bases 1 to 558)
JOURNAL Direct Submission
Submitted (02-JUL-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: CLOMTECH
CDNA Library Preparation: CLOMTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
Contact: amadan@systemsbiology.org
Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
Greene, Mark Keltman and Anuradha Madan

FEATURES
source location/Qualifiers
1..558 /organism="Homo sapiens"
/db_xref="locusID:3347"
/db_xref="taxon:9606"
/clone="MGC:13578 IMAGE:4293405"
/tissue_type="Skeletal Muscle"
/clone_lib="NIH-MGC_81"
/lab_host="DH10B"
/note="Vector: pDNR-LIB"
76..231
/codon_start=1
/product="histatin 3"
/protein_id="AAH09791.1"
/db_xref="GI:14602561"
/translation="MKFFVALLALMLSMTGADSHAKRHGKRRFHEKHSHRGYR
SNLYDN"
BASE COUNT 187 a 102 c 91 g 178 t
ORIGIN

Query Match 6.5%; Score 86.2; DB 9; Length 558;
Best Local Similarity 62.5%; Pred. No. 3.2e-08;
Matches 150; Conservative 0; Mismatches 89; Indels 1; Gaps 1;

QY 1093 CACACCATACCAACCAATACCAATATGTTTAATATCATCTAGTACGAG 1152
Db 191 CACATCGAGGCTATGATCAATATCTGTATGACATGATTCCTTCATATCATGGG 250
QY 1153 ACATGATTTATGAGGCTTGATGGCAATATGACATTCATATTCATCTTCA 1212
Db 251 GCATGATTTATGAGGCTTGATGGCAATATGACATTCATATTCATCTTCA 310
QY 1213 TACCATATCAGACTACTACCATCTTTGTGATGATCTCTTAAGCAATGC-GAATGTAA 1271
Db 311 TACCGGATCAGACTACTACCATCTTTGTGATGATCTCTTAAGCAATGC-AATGAAG 370
QY 1272 ACCCTAATATTTACTGATATCTTTGTTCCAGATATCTGCTTTCATGTCACCTG 1331
Db 371 AATATACATGATTTAGTGAATCTGTGTTTCAGATATCTCCCTTCCTAATATTCATTGG 430

RESULT 15
LOCUS HUMHIS1X 480 bp mRNA linear PRI 31-DEC-1994
DEFINITION Human histatin 1 (HIS1) mRNA, complete cds.
ACCESSION M26664
VERSION M26664.1 GI:292143

KEYWORDS	hista1n 1.
SOURCE	Homo sapiens parotid gland cDNA to mRNA
ORGANISM	Homo sapiens

SOURCE ORGANISM	CDNA to mRNA.
Homo sapiens	parotid gland
Homo sapiens	

ORGANISM	Homo sapiens
Pathogen	

REFERENCE
ARTHURS
1 (Pages 1 to 480)
Cebidae; Hominae; Homo.
Catarrhini; Primates;
Mammalia; Eutheria;
Chiroptera; Insectivora;
Euteleostomi; Vertebrata;
Mollusca; Metazoa;

NOV 1963
SADDLEHILL, L.M. AND A
HISTORICAL SOCIETY

histatins, a family of salivary histidine-rich proteins, are encoded by at least two loci (HIS1 and HIS2)

PI Tang YF, Corley NC, Guegler KJ, Patterson C;
 XX MPI: 2000-350699/30.
 DR P-PSDB: AAY94526.
 XX

PT Purified polypeptide used for treating or preventing a disorder
 PT characterized by expression or activity of lysine-rich statherin
 PT proteins -
 XX

PS Claim 3; Page 70; 75pp; English.

XX The present sequence is human lysine-rich statherin protein (LRSP)
 CC cDNA from Inocyte clone 2820214. This sequence was identified
 CC through analysis of a cDNA library of breast tumor tissue
 CC (BRSTN014). The LRSP sequence was found to have homology with
 CC human statherin (AAY94527) and human basic histidine-rich protein
 CC (AAY94528). Human statherin is a phosphoprotein that acts as an
 CC inhibitor of precipitation of calcium phosphate salts in the oral
 CC cavity. The LRSP polypeptide and its antagonists may be useful for
 CC treating or preventing disorders associated with the activity of
 CC LRSP. Such disorders include autoimmune/inflammatory disorders (for
 CC example AIDS, allergies, asthma, diabetes mellitus), bacterial and
 CC fungal infection and cancers (such as leukemia, adenocarcinoma,
 CC melanoma). Antibodies to LRSP may be useful for diagnosis of
 CC the above disorders.
 XX

SQ Sequence 1331 BP; 397 A; 258 C; 228 G; 447 T; 1 other;

Query Match 99.9%; Score 1330; DB 21; Length 1331;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 CTAATGTTTAAAGTAAAGATGACCGGTAGCTGTCTCAATGACCAAGCTGAATTT 60
DB 1 CTAATGTTTAAAGTAAAGATGACCGGTAGCTGTCTCAATGACCAAGCTGAATTT 60
QY 61 AAGAGCTTTACAAATGTCGGAATTTTGCATGCGCATAGGGAATGTAAGTTACTTGGC 120
DB 61 AAGAGCTTTACAAATGTCGGAATTTTGCATGCGCATAGGGAATGTAAGTTACTTGGC 120
QY 121 TGGAAATTTATCAGACTGTGAGTAAACAAGTGAAGTTACAGATGAGGGGAATATG 180
DB 121 TGGAAATTTATCAGACTGTGAGTAAACAAGTGAAGTTACAGATGAGGGGAATATG 180
QY 181 AGGCCCCCTAAGGCTAAACAATAATCAATCTAGATAGTGGTATGCGTCCCA 240
DB 181 AGGCCCCCTAAGGCTAAACAATAATCAATCTAGATAGTGGTATGCGTCCCA 240
QY 241 GGCCTAATTTGGGACACTTTTCTGATGCTTTGAGAGTACTTTCTTTTGAAGAAA 300
DB 241 GGCCTAATTTGGGACACTTTTCTGATGCTTTGAGAGTACTTTCTTTTGAAGAAA 300
QY 301 TTTTCATTTCTGCTGCGCATGCTATATCTCCCTTTATAGAGCATTTGATTTCTTCC 360
DB 301 TTTTCATTTCTGCTGCGCATGCTATATCTCCCTTTATAGAGCATTTGATTTCTTCC 360
QY 361 TTTTCTGGGAAATGTCCTATAGCATTTTGAATCTTTTGAATGCACTATGCAATTA 420
DB 361 TTTTCTGGGAAATGTCCTATAGCATTTTGAATCTTTTGAATGCACTATGCAATTA 420
QY 421 TGGTAATCCGCTTATGCTGAATACAGATGTAATTAACGCTTACATAATCTACA 480
DB 421 TGGTAATCCGCTTATGCTGAATACAGATGTAATTAACGCTTACATAATCTACA 480
QY 481 CTTGGATTTGCTGCACTCTACCAATAGCTTTTGAATGCACTGAAGTGTAAACAGAA 540
DB 481 CTTGGATTTGCTGCACTCTACCAATAGCTTTTGAATGCACTGAAGTGTAAACAGAA 540
QY 541 AGAGGCAATGCTGAGAAAGATAGCTAATATTTTGGTACTTTATCTGAATCAAG 600
DB 541 AGAGGCAATGCTGAGAAAGATAGCTAATATTTTGGTACTTTATCTGAATCAAG 600
QY 601 ATGCTGCTCCCTGACAGTGTGTTTCTTCTTACAGATCCTCATGTAATCCCTCGGA 660

```

```

DB 601 ATGCTGCTCCCTGACAGTGTGTTTCTTCTTACAGATCCTCATGTAATCCCTCGGA 660
QY 661 GCACAGGACAGTATAGTACACTCTCCATTTCTTTGTTTGTATTAAGACAGACTCT 720
DB 661 GCACAGGACAGTATAGTACACTCTCCATTTCTTTGTTTGTATTAAGACAGACTCT 720
QY 721 GTCTCAAAAAAAGACATTTATCATTAATACACTTTATAGAGCCCTAATTTCTTATC 780
DB 721 GTCTCAAAAAAAGACATTTATCATTAATACACTTTATAGAGCCCTAATTTCTTATC 780
QY 781 TGAAGCAGCTGTTTTTTTTTAAACAGTTAAGTACGATGTCACAGACAAATTTCTC 840
DB 781 TGAAGCAGCTGTTTTTTTTTAAACAGTTAAGTACGATGTCACAGACAAATTTCTC 840
QY 841 GATCAGATAGTCCCTGTCACAGTACCAATGTGTTTCATTAAGTGGGAAGAAACAG 900
DB 841 GATCAGATAGTCCCTGTCACAGTACCAATGTGTTTCATTAAGTGGGAAGAAACAG 900
QY 901 CATTTTAAAGTACTTTTGGGAGACTGATTTGAGTAATTAATTAACCTGCTCCCTT 960
DB 901 CATTTTAAAGTACTTTTGGGAGACTGATTTGAGTAATTAATTAACCTGCTCCCTT 960
QY 961 AAGAAAAAACCCTTCCACCTTACTGTGCTATTATATCCCTTACTTCCAAAGTTA 1020
DB 961 AAGAAAAAACCCTTCCACCTTACTGTGCTATTATATCCCTTACTTCCAAAGTTA 1020
QY 1021 ATTATCTTATTTGATGATTTGCTTTTATACCAAGAGCTTATCAGCCAGTTCCAGAAC 1080
DB 1021 ATTATCTTATTTGATGATTTGCTTTTATACCAAGAGCTTATCAGCCAGTTCCAGAAC 1080
QY 1081 AACCACTATACGACACACATACCAACCAACCAATACCAATATAGCTTTATATATC 1140
DB 1081 AACCACTATACGACACACATACCAACCAACCAATACCAATATAGCTTTATATATC 1140
QY 1141 AGTAACTGACAGGACATGATTTTGAAGCTTGAATGGAATATGCAATCTCTACATCATAT 1200
DB 1141 AGTAACTGACAGGACATGATTTTGAAGCTTGAATGGAATATGCAATCTCTACATCATAT 1200
QY 1201 TCTCATCTTTCATATACCATATACACACTACACTACACTTTTGTAGATCAATGAAGCAAT 1260
DB 1201 TCTCATCTTTCATATACCATATACACACTACACTACACTTTTGTAGATCAATGAAGCAAT 1260
QY 1261 GCGAATGTAAACCTATATATTTACTGATACTCTTTGTTCCAGATACTGCTTTTCC 1320
DB 1261 GCGAATGTAAACCTATATATTTACTGATACTCTTTGTTCCAGATACTGCTTTTCC 1320
QY 1321 AATGTCACCTG 1331
DB 1321 AATGTCACCTG 1331

RESULT 2
AAA48964
ID AAA48964 standard; DNA; 552 Bp.
AC AAA48964;
DT 06-OCT-2000 (first entry)
XX
DE Human statherin DNA.
XX
KW Lysine-rich statherin protein; LRSP; acidic phosphoprotein; human;
KW precipitation inhibitor; autoimmune; inflammatory disorder; AIDS;
KW asthma; allergy; diabetes mellitus; fungal; bacterial infection;
KW cancer; leukemia; adenocarcinoma; melanoma; ds.
XX
OS Homo sapiens.
XX
FH Key 73..261 Location/Qualifiers
FT CDS
FT /tag= a
FT /product= Statherin

```


QY 1022 TTTATTTTTCGATATGCTTTTATACCAAGGCTTATCCAGCTTCCAGANCA 1081
 Db 1578 TTTGCTTACATTTGGAGATGTTGGTATGGGCTTATATCCAGCTTCCAGANCA 1637
 QY 1082 ACCATATATGAGACCAATACCAACCAATACCAATATGTTTATATATCA 1141
 Db 1638 ACCATATATGAGACCAATACCAACCAATACCAATATGTTTATATCA 1697
 QY 1142 GTAACGACGACATGATTTATGAGGCTTATGCAATATGCAATGCTTATAT 1201
 Db 1698 GTAACGACGACATGATTTATGAGGCTTATGCAATATGCAATGCTTATAT 1757
 QY 1202 CTGATCTTTCATACATATACATACATCTTTCATGATGATGATGATG 1261
 Db 1758 CTGATCTTTCATACATATACATACATCTTTCATGATGATGATGATG 1817
 QY 1262 CGATATGTA---AAACCTATATATTTACTGATGATGATGATGATGATG 1318
 Db 1818 CAGATATTAAGAAATACATGATGATGATGATGATGATGATGATGATG 1877
 QY 1319 CCAATGTCATG 1331
 Db 1878 TAAATATATTTG 1890

RESULT 4

AAAA2355
 ID AAA42355 standard; cDNA: 221 BP.

AC AAA42355;

DT 21-AUG-2000 (first entry)

DE Human secreted expressed sequence tag SEQ ID NO:1095.

XX Human: mouse; xenopus; rat; secreted expressed sequence tag; SEST;
 KW expressed sequence tag; EST; probe; chemotactic; proliferative;
 KW immunomodulatory; haemopoietic; chemokine; analgesic; haemostatic;
 KW thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
 KW antiviral; antidiabetic; antisthmatic; neurotrophic; antiparkinsonian;
 KW antitumor; osteoprotective; neuroprotective; nootropic; antiparkinsonian;
 KW cerebroprotective; anticonvulsant; antidepressant; gene therapy;
 KW vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
 KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
 KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
 KW central nervous system disorder; Alzheimer's disease; stroke;
 KW Parkinson's disease; Huntington's disease; coagulation disorder;
 KW hemophilia; thrombosis; inflammatory disorder; Crohn's disease;
 KW tumour; infection; depression; psoriasis; ss.

OS Homo sapiens.

PN WO200021990-A1.

XX 20-APR-2000.

PE 15-OCT-1999; 99WO-US24205.

XX 15-OCT-1998; 98US-0104435.

XX (GENY) GENETICS INST INC.

PA Jacobs K, McCoy JM, Lavallee ER, Collins-Racle LA, Evans C;
 PI Merberg D, Treacy W;

XX WPI; 2000-317937/27.

DR Isolated polynucleotides, and encoded proteins, comprising secreted
 PT expressed sequence tags (SESTs), useful for treating various disorders
 PT such as autoimmune, infectious, and central nervous system disorders -

XX Claim 1; Page 394; 618pp; English.

PS

CC AAA4261 to AAA4349 represent specifically claimed secreted expressed
 CC sequence tags (SESTs), isolated from human, mouse, xenopus and rat
 CC tissue sources. The SESTs can have a range of activities depending on
 CC the tissues they were isolated from. The activities include:
 CC chemotactic; proliferative; immunomodulatory; haemopoietic;
 CC chemokine; analgesic; haemostatic; thrombolytic; antiinflammatory;
 CC cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
 CC antisthmatic; neurotrophic; osteoprotective; neuroprotective;
 CC nootropic; antiparkinsonian; antidepressant; gene therapy;
 CC anticonvulsant; and antidepressant. The SESTs can be used for gene
 CC therapy and in vaccines. The SESTs are useful as probes for the
 CC identification and isolation of full-length cDNAs and genomic DNA
 CC molecules which correspond to the SESTs. Proteins encoded by the SESTs
 CC are useful in assays for determining biological activity and raising
 CC antibodies. They may be useful for treatment of autoimmune disorders
 CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
 CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
 CC osteoporosis, osteoarthritis, central nervous system disorders
 CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
 CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
 CC disease), tumours, bacterial, fungal or viral infections, depression and
 CC psoriasis. AAA43420 to AAA43425 represent linker variants which are given
 CC in the exemplification of the present invention.

SQ Sequence 221 BP; 72 A; 41 C; 44 G; 64 T; 0 other;

Query Match 14.1%; Score 187.4; DB 21; Length 221;
 Best Local Similarity 99.5%; Pred. No. 3.6e-37;

Matches 188; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 793 TTTTATTTTAAACAGTTAAGTATGATGATGATGATGATGATGATGATG 852
 Db 28 TTTTATTTTAAACAGTTAAGTATGATGATGATGATGATGATGATGATG 87
 QY 853 CCTGTCAACAGTGAATGATGATGATGATGATGATGATGATGATGATG 912
 Db 88 CCTGTCAACAGTGAATGATGATGATGATGATGATGATGATGATGATG 147
 QY 913 ACTTTTGGGAGACTGATTTGATTAATTAATTAATTAATTAATTAATTA 972
 Db 148 ACTTTTGGGAGACTGATTTGATTAATTAATTAATTAATTAATTAATTA 207
 QY 973 CCTTCCAC 981
 Db 208 CCTTCCGC 216

RESULT 5

AI91635
 ID AI91635 standard; cDNA: 390 BP.

AC AI91635;

DT 06-NOV-2001 (first entry)

DE Human polynucleotide SEQ ID NO 11695.

XX Human: cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation; ss.

OS Homo sapiens.

PN WO200164835-A2.

XX 07-SEP-2001.

XX 26-FEB-2001; 2001WO-US04927.

XX 28-FEB-2000; 2000US-0515126.

XX 18-MAY-2000; 2000US-0577409.

PR

XX

PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;
 XX
 DR WPI: 2001-514838/56.
 DR P-PSDB: AA011704.
 XX
 PT Isolated nucleic acids and polypeptides, useful for preventing
 PT diagnosing and treating e.g. leukaemia, inflammation and immune
 PT disorders -
 XX
 PS Claim 1: SEQ ID NO 11695; 13999p + Sequence listing; English.
 XX
 CC The invention relates to human polynucleotides (AA17994-AA19341) and
 CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activity/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 SO Sequence 390 BP; 127 A; 77 C; 79 G; 106 T; 1 other;
 XX
 Query Match
 Best Local Similarity 6.5%; Score 86.6; DB 22; Length 390;
 Matches 122; Conservative 0; Mismatches 34; Indels 5; Gaps 1;
 XX
 OY 812 AGTACGATGTCACAGACAAATTTTCATGATAGTCCCTGTCACAGTACCAAA 871
 DB 152 AGAAAAATGATGAAATTTTCCAAATATTTCCGATCAGAGAAATC-----ACAAGACGACAAA 206
 OY 872 TGTGCTTTCATTAAGTGGGAAAGAAACAGCATTTTAAAGTAATTTTGGGAGACGAT 931
 DB 207 TGTGCTTTCATCAAGTGGGAAAGAAACAGCATTTTAAATTAATTTTGGGAGACGAT 266
 OY 932 TGAGTAAATTAATTAATCTGCTGCTCCCTTAAGAAAAAAA 972
 DB 267 TGAGTAAATTAATTAATCTGCTGCTCCCTTAATTAATTAATTA 307
 XX
 RESULT 6
 AAH98658
 ID AAH98658 standard; cDNA; 857 BP.
 XX
 AC AAH98658;
 XX
 DT 12-OCT-2001 (first entry)
 XX
 DE Human EST-derived coding sequence SEQ ID NO: 515.
 XX
 KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
 KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
 KW diagnostics; forensic test; gene mapping; genetic disorder;
 KW biodiversity; gene therapy; nutrition; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200154477-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 25-JAN-2001; 2001WO-US02687.
 XX
 PR 25-JAN-2000; 2000US-0491404.
 PR 17-JUL-2000; 2000US-0617746.
 PR 03-AUG-2000; 2000US-0631451.

PR 15-SEP-2000; 2000US-0663870.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
 PI Cao Y, Drmanac RA, Zhang J, Werhman T;
 XX
 DR WPI: 2001-476164/51.
 DR P-PSDB: AAM23999.
 XX
 PT Isolated polypeptide for treatment of diseases, diagnostics, raising
 PT antibodies and research use -
 XX
 PS Claim 1: Page 537; 1275pp; English.
 XX
 CC The present invention provides the protein and coding sequences of novel
 CC proteins from a variety of organisms, including human, dog, cat, horse,
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
 CC from the organism of interest. They can be used in diagnostics,
 CC forensics, gene mapping, identification of mutations, to assess
 CC biodiversity and for nutritional purposes. The present sequence is a cDNA
 CC of the invention.
 CC
 XX
 SO Sequence 857 BP; 231 A; 194 C; 169 G; 263 T; 0 other;
 XX
 Query Match
 Best Local Similarity 6.5%; Score 86.2; DB 22; Length 857;
 Matches 150; Conservative 0; Mismatches 89; Indels 1; Gaps 1;
 XX
 OY 1093 CACACCAATACCAACCAATACCAATATGCTTTAAATATCATCATGATCAGG 1152
 DB 510 CACATGAGGCTATATATCAATTAATCTGATGACATATGATCTTCAGTATCATGG 569
 OY 1153 ACATGATATTTGAGGCTGATGAGCAATATACGACTCTTCATCATATTTCTTCA 1212
 DB 570 GCATGATTTTGGAGGTTTACTGCGCAATTCCTTTGGACTCGTGTATTCATTTGTCA 629
 OY 1213 TACCATATACACTACTACTGTTTGTGATCATCTTAAGAGCAATGC-GAATGTAA 1271
 DB 630 TACCGCATCACACTACCACTGCTTTTGAAGATTTATCAATAGGCAATGCAGATTAAG 689
 OY 1272 ACCCTATATTTACTGATGATCTTTGTTCCAGATCTGCTTTTCCAATGCTCTTG 1331
 DB 690 AATACCATGATTTAGTGAATTCGTGTTTCAGATCTTCCCTTCCTAATATCATTTG 749
 XX
 RESULT 7
 AAA48965
 ID AAA48965 standard; DNA; 491 BP.
 XX
 AC AAA48965;
 XX
 DT 06-OCT-2000 (first entry)
 XX
 DE Human basic histidine-rich protein DNA.
 XX
 KW Lysine-rich statherin protein; LRSP; acidic phosphoprotein; human;
 KW precipitation inhibitor; autoimmune; inflammatory disorder; AIDS;
 KW asthma; allergy; diabetes mellitus; fungal; bacterial infection;
 KW cancer; leukemia; adenocarcinoma; melanoma; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key
 FH CDS Location/Qualifiers
 FT 38..193
 FT /*tag= a
 FT /product= Basic_histidine-rich_protein
 XX
 PN WO200024779-A1.
 XX
 PD 04-MAY-2000.

PF 22-OCT-1999; 99WO-US24046.
 XX
 PR 23-OCT-1999; 98US-0155209.
 XX
 PA (INCYTE) INCYTE PHARM INC.
 XX
 PI Tang YT, Corley NC, Guegler KJ, Patterson C;
 XX
 DR WPI: 2000-350699/30.
 XX
 DR P-PSDB: AAY94528.
 XX
 PT Purified polypeptide used for treating or preventing a disorder
 PT characterized by expression or activity of lysine-rich statherin
 PT proteins -
 PS
 PS Disclosure; Page 72; 75pp; English.
 CC The present invention relates to human lysine-rich statherin protein
 CC (LRSP) (AAY94526). The cDNA sequence encoding this protein was identified
 CC through analysis of a cDNA library of breast tumour tissue (BRSTM014).
 CC The LRSP sequence was found to have homology with human statherin
 CC protein (AAY94527) and human basic histidine-rich protein (the present
 CC sequence). Human statherin is a phosphoprotein that acts as an inhibitor
 CC of precipitation of calcium phosphate salts in the oral cavity. The LRSP
 CC polypeptide and its antagonists may be useful for treating or preventing
 CC disorders associated with the activity of LRSP. Such disorders include
 CC autoimmune/inflammatory disorders (for example AIDS, allergies, asthma,
 CC diabetes mellitus), bacterial and fungal infection and cancers (such as
 CC leukemia, adenocarcinoma, melanoma). Antibodies to LRSP may be useful
 CC for diagnosis of the above disorders.
 XX
 SO Sequence 491 BP; 151 A; 90 C; 87 G; 163 T; 0 other;
 Query Match 5.9%; Score 78.8; DB 21; Length 491;
 Best Local Similarity 60.6%; Pred. No. 6.5e-10;
 Matches 146; Conservative 0; Mismatches 93; Indels 2; Gaps 1;
 QY 1093 CACACACATACCAACCAATATACGTTTAAATATCATCAGTAAGTCAGG 1152
 DB 153 CACATCGAGCGCTATGATCAATATCTGTATGACATGATCTCTGATATCATGG 212
 QY 1153 ACATGATTTTATGAGCTTATGGAATATGACATCTTACATCATATTCATCTTCA 1212
 DB 213 GCATGATTTTATGAGGTTTATGAGTGCATATTCCTTGGACGTCTATTCATTTGCA 272
 QY 1213 TACCATATCAGTACTTACCTTTTGTGTTTGAAGATTTATCATTAAGCAATGAA 332
 DB 273 TACCGCATCAGTACTTACTGCTTTTGAAGATTTATCATTAAGCAATGAA 332
 QY 1271 AACCTATATATTTACTGATTAATCTTGGTTCCAGATTAATGCTTTTCAATGCACTT 1330
 DB 333 GAATATCATGATTAATGATTAATCTGTTTCAAGATTAATGCTTTTCAATGCACTT 1330
 QY 1331 G 1331
 DB 393 G 393
 RESULT 8
 AAT24320
 ID AAT24320 standard; cDNA to mRNA; 126 BP.
 XX
 AC AAT24320;
 XX
 DE 22-SEP-1996 (first entry).
 XX
 DE Human gene signature HUMGS06347.
 XX
 KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
 KW human; cloning; mapping; non-biased library; diagnosis; detection;
 KW cell typing; abnormal cell function; ss.
 XX
 OS Homo sapiens.

XX
 PN WO9514772-A1.
 XX
 PD 01-JUN-1995.
 XX
 PF 11-NOV-1994; 94WO-JP01916.
 XX
 PR 12-NOV-1993; 93JP-0355504.
 XX
 PA (MATSUBARA K.
 PA (OKUBO K.
 XX
 PI Matsubara K, Okubo K;
 XX
 DR WPI: 1995-206931/27.
 XX
 PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
 PT for diagnosis of abnormal cell function, by preparing cDNA that
 PT reflects relative abundance of corresp. mRNA in specific human
 PT tissues
 PS
 PS Claim 1; Page 1582; 2245pp; Japanese.
 CC A single-stranded DNA (or its complementary strand or the corresp.
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
 CC given in AAT19001-726837 and which is able to hybridize to part of
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
 CC sequences were obtained from 3'-directed cDNA libraries prepared
 CC from various human tissues; synthesis of cDNA was initiated from the
 CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
 CC untranslated sequence is unique to a particular mRNA species, almost
 CC all the 3'-oriented cDNAs hybridize with specific mRNA species. Each library
 CC is constructed so as to reflect accurately the relative abundance of
 CC different mRNAs in the particular tissue from which it was derived.
 CC The appearance frequency of a given GS in a cDNA library can be
 CC determined (esp. using primers and probes derived from the GS
 CC sequences) as a means of diagnosing abnormal cell function or for
 CC recognising different cell types.
 XX
 SO Sequence 126 BP; 44 A; 16 C; 25 G; 34 T; 7 other;
 Query Match 5.9%; Score 78.6; DB 16; Length 126;
 Best Local Similarity 86.6%; Pred. No. 4.6e-10;
 Matches 84; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
 QY 859 CAACAGTACCAATATGTTTCAATAGTGGGAAGAAACGATTTAACTACTTT 918
 DB 14 CAAGAGCAGCAATATGTTTCAATAGTGGGAAGAAACGATTTAACTACTTT 73
 QY 919 TGGGAGACGATTTGAGTAATATAAATCACTGCTCT 955
 DB 74 TGGGAGACGATTTGAGTAATATAAATCACTGCTCT 110
 RESULT 9
 AAT23140
 ID AAT23140 standard; cDNA to mRNA; 90 BP.
 XX
 AC AAT23140;
 XX
 DE 23-AUG-1996 (first entry)
 XX
 DE Human gene signature HUMGS04875.
 XX
 KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
 KW human; cloning; mapping; non-biased library; diagnosis; detection;
 KW cell typing; abnormal cell function; ss.
 XX
 OS Homo sapiens.
 PN WO9514772-A1.
 XX
 PD 01-JUN-1995.

XX 11-NOV-1994; 94WO-JP01916.
 PF 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX (EPIC-) EPIDEMIOLOGICS AG.
 PA Olek A, Piepenbrock C, Berlin K;
 PI WPI; 2002-017444/02.
 DR
 XX
 PT Chemically modified sequences of genes associated with apoptosis are
 PT useful to determine methylation patterns of genomic DNA samples for
 PT diagnosis of associated diseases such as cancer
 PS Claim 1; Seq ID #53; 24pp; English.
 XX
 PS Claim 1; Page 1297; 2245pp; Japanese.
 CC A single-stranded DNA (or its complementary strand or the corresp.
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
 CC given in AAT19001-726837 and which is able to hybridise to part of
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
 CC sequences were obtained from 3'-directed cDNA libraries prepared
 CC from various human tissues; synthesis of cDNA was initiated from the
 CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
 CC untranslated sequence is unique to a particular mRNA species, almost
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
 CC is constructed so as to reflect accurately the relative abundance of
 CC the appearance frequency of a given GS in a cDNA library can be
 CC determined (esp. using primers and probes derived from the GS
 CC sequences) as a means of diagnosing abnormal cell function or for
 CC recognising different cell types.
 SQ Sequence 90 BP; 30 A; 12 C; 17 G; 31 T; 0 other;
 XX
 Query Match 5.7%; Score 76.4; DB 16; Length 90;
 Best Local Similarity 97.8%; Pred. No. 1.4e-09;
 Matches 88; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
 QY 392 GATCTTTGATGCGCACTAATGCAATATGTAATGCC-GTATTGGCAATACAGCAT 450
 DB 1 GATCTTTGATGCGCACTAATGCAATATGTAATGCCGTTATGCGAATACAGCAT 60
 QY 451 AGTTAATAAAGCTGTACAGTAATCTACA 480
 DB 61 AGTTAATAAAGCTGTACAGTAATCTACA 90
 RESULT 10
 ID ABL54353 standard; DNA; 9881 BP.
 XX
 AC ABL54353;
 XX
 DT 29-JUL-2002 (first entry)
 XX
 DE Chemically treated apoptosis gene #27.
 XX
 KW Apoptosis; HIV; Bloom syndrome; cardiopathy;
 KW neurodegenerative disorder; Herpes simplex virus; renal ischaemia;
 KW amyotrophic lateral sclerosis; cancer; ds.
 OS Unidentified.
 XX
 PN WO200177164-A2.
 XX
 PD 18-OCT-2001.
 XX
 PF 06-APR-2001; 2001WO-EP03969.
 XX
 PR 06-APR-2000; 2000DE-1019058.

PR 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX (EPIC-) EPIDEMIOLOGICS AG.
 PA Olek A, Piepenbrock C, Berlin K;
 PI WPI; 2002-017444/02.
 DR
 XX
 PT Chemically modified sequences of genes associated with apoptosis are
 PT useful to determine methylation patterns of genomic DNA samples for
 PT diagnosis of associated diseases such as cancer
 PS Claim 1; Seq ID #53; 24pp; English.
 XX
 PS This invention relates to chemically pre-treated DNA of genes
 CC associated with apoptosis. The nucleic acids are used to allocate
 CC patients for specific therapy for HIV infection, Bloom syndrome,
 CC cardiopathy, aging, neurodegenerative disorders, Herpes simplex virus
 CC infection, renal ischaemia, amyotrophic lateral sclerosis, solid tumours
 CC and cancers. This nucleotide sequence represents a chemically
 CC treated apoptosis gene. Even SEQ ID numbers are the complementary
 CC DNA strands to the odd SEQ ID numbers. The sequence data for this
 CC patent is not represented in the printed specification but is based on
 CC information supplied by the European patent office.
 SQ Sequence 9881 BP; 2606 A; 196 C; 2308 G; 4771 T; 0 other;
 XX
 Query Match 5.6%; Score 74.6; DB 24; Length 9881;
 Best Local Similarity 78.8%; Pred. No. 2e-08;
 Matches 89; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
 QY 860 AACAGTAGCAAAATGCTTTCATTAAGTGGGAAGAAACGATTTTAAAGTACTTTT 919
 DB 5771 AACAGTAGCAAAATGCTTTCATTAAGTGGGAAGAAAGTAAATTAATATTTT 5830
 QY 920 GGGAGCGCATTTTGAATTAATTAACCTGCTCCCTTGAAGAAAAAAA 972
 DB 5831 GGGAGCTGAATTTGAATTAATTAATTAATTTTGTGTTAATTAATTAATTA 5883
 RESULT 11
 ID AAV89525 standard; cDNA; 516 BP.
 XX
 AC AAV89525;
 XX
 DT 15-FEB-1999 (first entry)
 XX
 DE EST clone CP289.
 XX
 KW Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
 KW tissue growth; activin; inhibin; chemotaxis; chemokinesis; hemostatic;
 KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
 KW gene therapy; ss.
 OS Homo sapiens.
 XX
 PN WO9845436-A2.
 XX
 PD 15-OCT-1998.
 XX
 PF 10-APR-1998; 98WO-US06955.
 XX
 PR 10-APR-1997; 97US-0838821.
 XX
 PA (GEMV) GENETICS INST INC.
 XX
 PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
 PI Racie LA, Spaulding V, Treacy M;
 DR WPI; 1999-070077/06.

XX New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries.

PS Claim 1: Page 245: 618pp: English.

XX The present sequence represents a human expressed sequence tag (EST).
CC The polynucleotide, which is a secreted EST, and the encoded protein
CC are predicted to have useful biological activities which would make
CC them suitable for treating, preventing or ameliorating medical
CC conditions in humans and animals, although no supporting data is
CC given. Suggested activities include nutritional activity, immune
CC stimulating or suppressing activity, haematopoiesis regulating
CC activity, tissue growth activity, activin/inhibin activity
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC activity, receptor/ligand activity, anti-inflammatory activity,
CC cachectin/tumour invasion suppressor activity, tumour inhibition
CC activity. The polynucleotide may also be useful for gene therapy.

XX Sequence 516 BP; 156 A; 181 C; 97 G; 82 T; 0 other:

Query Match 4.5%: Score 60.2; DB 20; Length 516;
Best Local Similarity 89.0%: Pred. No. 2.9e-05;
Matches 65; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 1047 TATACCAAGAGCCCTATACGACGATTCAGACACCACTATACGACACCAATACCA 1106
DB 139 TATGGGATATGCGCCCTATACGACGATTCAGACACCACTATACGACACCAATACCA 198
OY 1107 CCACATATACCAAC 1119
DB 199 CCACATATACCAAC 211

RESULT 12
ABN43273
ID ABN43273 standard; DNA: 60 BP.

XX ABN43273:
AC 15-JUL-2002 (first entry)

XX Human spliced transcript detection oligonucleotide SEQ ID NO:16021.
DE Human spliced transcript detection oligonucleotide; detection: RNA transcript;
XX splice variant; transcriptome; oligonucleotide library; ss.

XX Homo sapiens.
OS Homo sapiens.
XX WO200210449-A2.
XX PN 07-FEB-2002.
XX PD 20-JUL-2001; 2001WO-1B01903.
XX PF 28-JUL-2000; 2000US-221607P.
XX PR 02-MAY-2001; 2001US-287724P.
XX XX (COMP-) COMPUGEN INC.

XX Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
PI WPI: 2002-257383/30.
XX DR 2002-017444/02.

XX New oligonucleotide libraries comprising oligonucleotides which
PT selectively hybridize to mRNAs transcribed from a transcription unit of
PT a genome, useful for detecting tissue-, pathology-, and
PT developmental-specific genes -
XX Example 1: SEQ ID 16021: 47pp: English.
XX The present invention describes oligonucleotide libraries for detecting

CC messenger RNAs that populate a (sub-)transcriptome, where the
CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple
CC transcription units that populate a genome. The library comprises
CC several oligonucleotides, each capable of hybridizing selectively to a
CC set of messenger RNAs transcribed from a given transcription unit of
CC the genome, which encodes one or more messenger RNA splice variants.
CC The oligonucleotide libraries are useful for detecting mRNAs from a
CC biologically sample, in expression profiling studies, in qualitatively or
CC quantitatively characterizing the corresponding transcriptome, and in
CC detecting RNA transcripts and splice variants of human or animal
CC transcriptomes. The libraries may also be used as specialized mini-
CC libraries to detect transcripts of a sub-transcriptome under a
CC particular biological or pathological state, and so allowing the
CC detection of tissue- and pathology-specific genes such as those genes
CC only expressed in specific tissue under a specific pathological
CC condition; to detect developmental specific genes; and to detect RNA
CC transcripts and splice variants of a transcriptome of a patient suffering
CC from a particular disorder. ABN27253 to ABN59589 represent
CC oligonucleotide sequences from rats, humans and mice, which are used in
CC the exemplification of the present invention.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 60 BP; 22 A; 25 C; 4 G; 9 T; 0 other:
Query Match 4.2%: Score 56.4; DB 24; Length 60;
Best Local Similarity 98.3%: Pred. No. 0.00012;
Matches 57; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1059 CCTATACGACGATTCAGACACCACTATACGACACCAATACCAATACCA 1116
DB 3 CCTATACGACGATTCAGACACCACTATACGACACCAATACCAATACCAATACCA 60

RESULT 13
ABL54354/C
ID ABL54354 standard; DNA: 9881 BP.

XX ABL54354:
AC 29-JUL-2002 (first entry)

XX Chemically treated apoptosis gene complementary to gene #27.

XX Apoptosis; HIV; Bloom syndrome; cardiopathy;
XX neurodegenerative disorder; Herpes simplex virus; renal ischaemia;
XX amyotrophic lateral sclerosis; cancer; ds.

XX Unidentified.
XX OS Unidentified.
XX PN WO200177164-A2.
XX PD 18-OCT-2001.
XX PF 06-APR-2001; 2001WO-EP03969.
XX PF 06-APR-2000; 2000DE-1019058.
XX PR 07-APR-2000; 2000DE-1019173.
XX PR 30-JUN-2000; 2000DE-1032529.
XX PR 01-SEP-2000; 2000DE-1043826.
XX XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;
PI WPI: 2002-017444/02.
XX DR 2002-017444/02.

XX Chemically modified sequences of genes associated with apoptosis are
PT useful to determine methylation patterns of genomic DNA samples for
PT diagnosis of associated diseases such as cancer -
XX Claim 1: Seq ID #54: 24pp: English.

XX This invention relates to chemically pre-treated DNA of genes
 CC associated with apoptosis. The nucleic acids are used to allocate
 CC patients for specific therapy for HIV infection, Bloom syndrome,
 CC cataplexy, aging, neurodegenerative disorders, Herpes simplex virus
 CC infection, renal ischaemia, amyotrophic lateral sclerosis, solid tumours
 CC and cancers. This nucleotide sequence represents a chemically
 CC treated apoptosis gene. Even SEQ ID numbers are the complementary
 CC DNA strands to the odd SEQ ID numbers. The sequence data for this
 CC patent is not represented by the printed specification but is based on
 CC information supplied by the European patent office.

XX Sequence 9881 BP; 2874 A; 196 C; 2093 G; 4718 T; 0 other;

Query Match 4.2%; Score 55.8; DB 24; Length 9881;
 Best Local Similarity 67.1%; Pred. No. 0.001;

Matches 96; Conservative 0; Mismatches 42; Indels 5; Gaps 1;

OY 830 CAATATATTCGATCAGATAGTCCCTGTCACAGTCAATGCTTCTAATAAGTGG 889
 DB 4136 CAATATATTCGATCAGATAGTCCCTGTCACAGTCAATGCTTCTAATAAGTGG 4082
 OY 890 GAAGAAAACAGCATTTTAAAGTACTTTTGGGAGACTGATTGAGTATATAAACTC 949
 DB 4081 AAAAAAACAACATTTAAATTAATTTTAAAAAATAAATAATAATAAACTT 4022
 OY 950 TGTCTCCCTTAAGAAAAA 972
 DB 4021 CAATCTTCGCTAATATATATTA 3999

RESULT 14

ID ABL33026 standard; DNA: 6056 BP.

XX ABL33026;

XX 26-MAR-2002 (first entry)

XX Human immune system associated gene SEQ ID NO: 999.

XX Human; immune system disease; cytosine methylation; antileukemic;
 KM antiarteriosclerotic; antianemic; cytosolic; neutrotrophic;
 KM neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KM antineuritic; antidiabetic; antiparasitic;
 KM antineuritic; cancer; eye disease; arteriosclerosis; anaemia;
 KM acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KM neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KM gene; ds.

XX Homo sapiens.

XX WO200200928-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX MPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation
 PS c1a1m 1; SEQ ID NO 999; 32pp + Sequence Listing; German.
 XX

CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.

XX Sequence 6056 BP; 1780 A; 54 C; 1079 G; 3143 T; 0 other;

Query Match 3.6%; Score 48.4; DB 24; Length 6056;
 Best Local Similarity 45.9%; Pred. No. 0.06;

Matches 166; Conservative 0; Mismatches 196; Indels 0; Gaps 0;

OY 688 TTCTCTTTTGGTTTAAAGACGACCTGCTCAAAAAAAGACATTTATCAT 747
 DB 3741 TTATATGATTTTTCGAAATGAAATTTTATTTTGTATTAATAATTAATAA 3800
 OY 748 ATACATCTTATTTAGAGCCCTAATTTCTATCTGAAGCAGCTTTTAAACA 807
 DB 3801 AGTAGTGTATTTATGTAAGATTTTGTATTTTGGAAATTTTGTATTAATA 3860
 OY 808 GTTAAGTACGATGTCACAGACAAATTTCTGATCAGATAGTCCCTGTCACAGTAG 867
 DB 3861 ATTTTGTATTTGCTAAATATTAATAATTAATAATTTTGTATTTTATTAATA 3920
 OY 868 CAATGTGTTCTCAATAAGTGGGAAAGAAACAGATTTTAAAGTACTTTTGGAGACT 927
 DB 3921 GTATATATGTTTACTATATGTTTAAAGAAAAATGTTAAAGAAATTTGTTTATTAAT 3980
 OY 928 GATTGTGATTAATAAATTAACCTGCTCCCTTAAGAAAAAACCCTCCACCTTTAC 987
 DB 3981 TATTTTAAAGATTTGATTAATTTATTTTGTATTTTAAAGCTTTATTTATAGATTAT 4040
 OY 988 TGTGCTATTTATATCCCTTAGTTTCAAGTTAATTAATTTTGTGATATTTGCTTTT 1047
 DB 4041 TTTTATATGTTGAATATATAGTTTAAATATATGATTTTAAATTTTAAATAT 4100
 OY 1048 AT 1049
 DB 4101 TT 4102

RESULT 15

ID AAS46787/c standard; DNA: 61020 BP.

XX AAS46787;

XX 18-DEC-2001 (first entry)

XX Tumour suppressor gene derived chemically modified sequence #513.

XX Human; tumour suppressor gene; oncogene; antitumour; cytosolic;
 KM cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
 KM cytosine methylation; ds.

XX Homo sapiens.

XX WO200168912-A2.

XX 20-SEP-2001.

XX 15-MAR-2001; 2001WO-EP02955.

XX 15-MAR-2000; 2000DE-1013847.

XX 06-APR-2000; 2000DE-1019058.

XX 07-APR-2000; 2000DE-1019173.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIGENOMICS AG.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2003, 14:15:20 ; Search time 92 Seconds

(without alignments)
4436.813 Million cell updates/sec

Title: US-09-830-244B-2

Perfect score: 1331

Sequence: 1 ctatgttttagaacaag.....gcctttccatgactctg 1331

Scoring table:

IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/1/1na/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/1na/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/1na/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/1na/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/1na/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/1/1na/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	39.6	3.0	7218	1 US-08-232-463-14	Sequence 14, Appl
2	38.6	2.9	6265	4 US-09-129-112-3	Sequence 3, Appl
3	38	2.9	588	1 US-09-385-982-133	Sequence 133, App
4	37.2	2.8	5852	1 US-07-867-106-2	Sequence 2, Appl
5	37.2	2.8	246240	2 US-08-724-394A-20	Sequence 20, Appl
6	37.2	2.8	246240	2 US-08-724-394A-21	Sequence 21, Appl
7	37.2	2.8	246240	2 US-08-724-394A-22	Sequence 22, Appl
8	36.4	2.7	12730	4 US-09-004-838-91	Sequence 91, Appl
9	36.4	2.7	12730	4 US-09-004-838-124	Sequence 124, App
10	36.2	2.7	762	4 US-09-134-001C-1748	Sequence 748, App
11	36.2	2.7	8302	4 US-09-234-827B-1	Sequence 1, Appl
12	36	2.7	3718	4 US-09-424-283-6	Sequence 6, Appl
13	35.8	2.7	72928	3 US-09-009-913-1	Sequence 1, Appl
14	35.6	2.7	3780	4 US-09-134-001C-990	Sequence 920, App
15	35.6	2.7	19124	2 US-08-487-826B-13	Sequence 13, Appl
16	35.6	2.6	1474	4 US-08-821-994-64	Sequence 14, Appl
17	35	2.6	2192	3 US-08-714-918-14	Sequence 14, Appl
18	35	2.6	2192	4 US-09-265-315-14	Sequence 14, Appl
19	35	2.6	2192	4 US-09-265-315-14	Sequence 14, Appl
20	35	2.6	2192	4 US-09-265-315-14	Sequence 14, Appl
21	35	2.6	9048	3 US-08-973-273-4	Sequence 4, Appl
22	35	2.6	36651	4 US-09-738-894A-3	Sequence 3, Appl
23	34.8	2.6	4817	4 US-09-140-804-9	Sequence 9, Appl
24	34.8	2.6	11056	4 US-09-004-838-23	Sequence 23, Appl
25	34.8	2.6	15062	4 US-09-004-838-89	Sequence 89, Appl
26	34.8	2.6	62804	4 US-09-800-960-3	Sequence 3, Appl
27	34.4	2.6	2173	6 5168051-10	Patent No. 5168051

28	34.2	2.6	688	4 US-08-998-416-915	Sequence 915, App
29	34.2	2.6	1394	4 US-09-247-155-76	Sequence 76, Appl
30	34.2	2.6	1408	3 US-08-889-841B-22	Sequence 22, Appl
31	34.2	2.6	7228	2 US-08-850-049-128	Sequence 128, App
32	34.2	2.6	7228	2 US-08-850-049-129	Sequence 129, App
33	34.2	2.6	7228	2 US-08-050-478-128	Sequence 128, App
34	34.2	2.6	7228	2 US-08-050-478-129	Sequence 128, App
35	34.2	2.6	7228	4 US-09-414-117-128	Sequence 128, App
36	34.2	2.6	7228	4 US-09-414-117-129	Sequence 129, App
37	34.2	2.6	7228	4 US-09-678-437-128	Sequence 128, App
38	34.2	2.6	7228	4 US-09-678-437-129	Sequence 129, App
39	34.2	2.6	15581	3 US-08-646-538-35	Sequence 35, Appl
40	34.2	2.6	15581	4 US-09-503-222-35	Sequence 35, Appl
41	34	2.6	1744	4 US-09-511-625B-19	Sequence 19, Appl
42	34	2.6	17425	4 US-09-511-625B-5	Sequence 5, Appl
43	34	2.6	32768	4 US-08-961-527-71	Sequence 71, Appl
44	33.8	2.5	1786	1 US-07-920-430-19	Sequence 19, Appl
45	33.8	2.5	1786	1 US-08-066-299-9	Sequence 9, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER-READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-Fls
US-08-232-463-14
Query Match 3.0%, Score 39.6, DB 1, Length 7218;

Best Local Similarity 7.0%; Pred. No. 0.54;
Matches 18; Conservative 138; Mismatches 102; Indels 0; Gaps 0;

QY 545 GCATGCTCTCAGAAAGATACCTAATATTTTGGTACTTATCTGAATCCAGATGC 604
DB 1037 GCTTGCTCGACATGCGAGCTGCGATATTTTTTTTTTTTTTTTTTTTTT 1096
QY 605 TCGTCCCTGAGGTTTCTTCTTACATCCTCTAATCCCTCTGGAGCAGC 664
DB 1097 TT 1156
QY 665 AGCAGATAGTACAGATCTCCATCTTCTTGTGTTTAAAGACAGATCTGCT 724
DB 1157 TT 1216
QY 725 CAAAAAAGACATTTATATATATATATATAGAGCCCTAATTTCTTATCTGAA 784
DB 1217 TT 1276
QY 785 GGCACCTGTTTTTTTTTTT 802
DB 1277 TTTTTTTTTTTTTTTTTT 1294

RESULT 2

US-09-129-112-3/c

; Sequence 3, Application US/09129112

; Patent No. 6465716

; GENERAL INFORMATION:

; APPLICANT: Etzler, Marilyn E.

; APPLICANT: Murphy, Judith B.

; TITLE OF INVENTION: The Regents of the University of California

; FILE REFERENCE: 023070-079810DS Factor Binding Protein From Legume Roots

; CURRENT APPLICATION NUMBER: US/09/129,112

; PRIOR FILING DATE: 1998-08-04

; PRIOR FILING DATE: US 08/907,226

; NUMBER OF SEQ ID NOS: 19

; SOFTWARE: Patent Ver. 2.1

; SEQ ID NO 3

; LENGTH: 6265

; TYPE: DNA

; ORGANISM: Dolichos biflorus

; FEATURE:

; OTHER INFORMATION: genomic sequence of NBP46 (DB46)

; NAME/KEY: exon

; LOCATION: (633)..(944)

; NAME/KEY: Intron

; LOCATION: (945)..(1022)

; NAME/KEY: exon

; LOCATION: (1023)..(1151)

; NAME/KEY: Intron

; LOCATION: (1152)..(1559)

; NAME/KEY: exon

; LOCATION: (1560)..(1616)

; NAME/KEY: Intron

; LOCATION: (1617)..(1697)

; NAME/KEY: exon

; LOCATION: (1698)..(1790)

; US-09-129-112-3

; Query Match

; Best Local Similarity 42.5%; Pred. No. 0.94;

; Matches 206; Conservative 0; Mismatches 279; Indels 0; Gaps 0;

QY 726 AAAAAAAGACATTTATATATATATATATAGAGCCCTAATTTCTTATCTGAG 785
DB 2790 AACTGAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2731
QY 786 GCATGCTTTTTTTTTTAAACAGTTAGTACTGATGTCACAGCAATATTTCTGATCA 845
DB 2730 ACTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2671

QY 846 GATAGTCCCTGTCACAGATAGCAATGTGTTTCATTAAGTGGAGAGAAACACATTT 905

DB 2670 AAAATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2611

QY 906 TAAAGTACTTTTTGGGAGACTGATTTGATTAATTAATTAATTAATTAATTAATTA 965

DB 2610 TTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2551

QY 966 AAAAAACCTTCCACCTTCTGTCATTTATATCCCTAGTCCCAAGTTAATAT 1025

DB 2550 CAAAGAACCTTAATTTCTTTTGAACATATACCTCATGTTTATCCACTTTAATTTTC 2491

QY 1026 CTATTTCTGATATATGCTTTTATATACCAAGAGCCCTATGAGCAGTCCAGAACCA 1085

DB 2490 ATCATTTGCTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 2431

QY 1086 CTATACGACACATACCATACCAACATACCAATATACGTTTAAATATATCATAGTAA 1145

DB 2430 GCGGACCTTAAGAAACACAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2371

QY 1146 CTGACGACATGATTTATGAGCTTGATGCAATACGACTTGCATCCATATTCATCA 1205

DB 2370 AATTAATATCTGCTGATTCATTCGACATTAATTAATTAATTAATTAATTAATTAAT 2311

QY 1206 TCTTT 1210

DB 2310 CATTT 2306

RESULT 3

US-09-385-982-133

; Sequence 133, Application US/09385982

; Patent No. 6262334

; GENERAL INFORMATION:

; APPLICANT: ENDEGE, WILLSON O., ET AL.

; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION

; FILE REFERENCE: CCDNA-260XX

; CURRENT APPLICATION NUMBER: US/09/385,982

; PRIOR FILING DATE: 1999-08-30

; PRIOR FILING DATE: 09/328,111

; EARLIER APPLICATION NUMBER: 60/117,393

; EARLIER FILING DATE: 1999-01-27

; EARLIER APPLICATION NUMBER: 60/098,639

; EARLIER FILING DATE: 1998-08-31

; NUMBER OF SEQ ID NOS: 544

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 133

; LENGTH: 388

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (1)..(588)

; OTHER INFORMATION: n = A,T,C or G

; US-09-385-982-133

; Query Match

; Best Local Similarity 41.3%; Pred. No. 0.57;

; Matches 95; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 35 TGTCATGTACCAACGTAAGTAATTTACAGTGTGTACCAATGTCTGGAATTTTGCACCTGC 94
DB 96 TGGCGNAGGTAACAGAGGTCAGATATGTCNNGTGACAGAGACAGCAANTCTGGCCNCA 155
QY 95 CATAGGAATGTTAAGGTTACTTGGCTGGAATTTATCAAGCTTGTGAGTAACAGTTGA 154
DB 156 CATTGANGTANNAANAAGTTTATNTTTTACANTTATNNNANNTATNNNANNTTTTAA 215
QY 155 AGTTAGCAGATGAGGGGAATTTAGAGCCCTTAAGCTTAACAAATTAATGATATCT 214

Db 216 NCTGCANCANNTGATTTTTCACACCTAANTTACTAGAAAACTAANGAAAAGCACTNATTAAGCT 2175

Qy 215 GAGATAGTGGCTAATGTGGCTCCCGCAGCTTAATTTGGAAACAGTTTTC 264

Db 276 CTGAATTAAANTTACATGNAAGCCTTTTCCTAATCTTNCANAAAACCTTC 325

RESULT 4
US-07-86

US-07-867-106-2
Sequence 2, Application US/07867106
Patent No. 5389526
GENERAL INFORMATION:
APPLICANT: Slade, Martin B
APPLICANT: Chang, Andy C M
APPLICANT: Williams, Keith L
TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
TITLE OF INVENTION: Slime Moulds of the Genus Dictyostellium
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526/ris
STREET: One Liberty Place 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103

COMPUTER READABLE FORM

```

1      MEDIUM TYPE: floppy disk
2      COMPUTER: IBM PC compatible
3      OPERATING SYSTEM: PC-DOS/MS-DOS
4      SOFTWARE: Patentm Release #1.0, Version #1.2
5      CURRENT APPLICATION DATA:
6      APPLICATION NUMBER: US/07/867,106
7      FILING DATE: 19920625
8      PRIOR APPLICATION DATA:
9      APPLICATION NUMBER: AU PJ 7187
10     APPLICATION NUMBER: PCT/AU90/00530
11     FILING DATE: 02-NOV-1989
12     ATTORNEY/AGENT INFORMATION:
13     NAME: Feeney, Joanne Longo
14     REGISTRATION NUMBER: 35,114
15     REFERENCE/DOCKET NUMBER: RICE-0002
16     TELECOMMUNICATION INFORMATION:
17     TELEPHONE: 215-568-3100
18     TELEFAX: 215-568-3439
19     INFORMATION FOR SEQ ID NO: 2:
20     SEQUENCE CHARACTERISTICS:
21     LENGTH: 5852 base pairs
22     TYPE: NUCLEIC ACID
23     STRANDEDNESS: single
24     TOPOLOGY: linear
25     MOLECULE TYPE: DNA (genomic)
26     ANTI-SENSE: NO
27     FEATURE:
28     NAME/KEY: CDS
29     LOCATION: 2378..5038
30     FEATURE:
31     NAME/KEY: CDS
32     LOCATION: 2378..5038
33     US-07-867-106-2

```

Query Match	2.88;	Score 37.2;	DB 1;	Length 5852;
Best Local Similarity	52.68;	Pred. No. 2.2;		
Matches 81;	Conservative 0;	Mismatches 73;	Indels 0;	Gaps 0

OY	682	TCTCCATTTCTTGGTTTTGTTTTTAAGACAGAGCCTGTCTCAAAAAAAGGCATT	741
Dd	1973	TTTGATTTCTTTTTTTTTTTTTTTTTTTTTTTTACTTGAAAAAAAATAAAAAAA	2033
OY	742	ATCATTAACAATCTATTAGAGCCCTAAATTTCTTATCGAAGCACGTGTTTTTTT	801
Dd	2033	AAAAAAAAAACCCCATTAATTAATATTAAATTAATTAATTAATTAATTAATTAATTA	2092
OY	802	TAAACAGTTAAGTACTGATGTCAACAGACAATA	835

Db 2093 ATAAATTAAATTTTATTCCTATCTAATTATA 2126

US-08-724-394A-20

Sequence 20, Application US/08724394A
Patent No. 5872237

GENERAL INFORMATION:

APPLICANT: Feder, John N.
APPLICANT: Krommal, Gregory S.
APPLICANT: Lauer, Peter M.
APPLICANT: Ruddy, David A.
APPLICANT: Thomas, Winston
APPLICANT: Tsuchinashi, Zenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237a1
TITLE OF INVENTION: Sequences and Antibodies Thereo
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:

STREET: Two Embarca

```

1      CITY: San Francisco
2      STATE: CA
3      COUNTRY: USA
4      ZIP: 94111-3634
5
6      COMPUTER READABLE FORM:
7      MEDIUM TYPE: Floppy disk
8      COMPUTER: IBM PC compatible
9      OPERATING SYSTEM: PC-DOS/MS-DO5
10     SOFTWARE: Patent Release #1.0, Version #1.3
11     CURRENT APPLICATION DATA:
12     APPLICATION NUMBER: US/08/724,394A
13     FILING DATE: 01-OCT-1996
14     CLASSIFICATION: 536
15
16     ATTORNEY/AGENT INFORMATION:
17     NAME: Filts, Renee A.
18     REGISTRATION NUMBER: 35,136
19     REFERENCE/DOCKET NUMBER: 017957-000100
20     TELECOMMUNICATION INFORMATION:
21     TELEPHONE: 415-576-0200
22     TELEFAX: 415-576-0300
23
24     INFORMATION FOR SEQ ID NO: 20:
25     SEQUENCE CHARACTERISTICS:
26     LENGTH: 246240 base pairs
27     TYPE: nucleic acid
28     STRANDEDNESS: not relevant
29     TOPOLOGY: not relevant
30
31     MOLECULE TYPE: cDNA
32
33     FEATURE:
34     NAME/KEY: misc.feature
35     LOCATION: 1..246240
36     OTHER INFORMATION: /note="HLA-H.CONFIG"
37
38     US-08-724-394A-20

```

Query Match	2.88;	Score 37.2;	DB 2;	length 246240;
Best Local Similarity	84.08;	Pred. No. 8.5;		
Matches 42;	Conservative 0;	Mismatches 8;	Indels 0;	Gaps 0

QY	912	AACCTTTTGGGAGACTGATTTGAGTAATAATAAACCTCGTGTCGCCCTTA	961
Db	154422	AATTTCGAAGGAGACTGATTTGAGTAATAATAAAACCTAGTCTCCCTGTA	154477

RESULT 6
US-08-724-394A-21

Sequence 21, Application US/08724339A
Patent No. 5872237
GENERAL INFORMATION:
APPLICANT: Feder, John N.
APPLICANT: Krommal, Gregory S.
APPLICANT: Laufer, Peter M.
APPLICANT: Ruddy, David A.

APPLICANT: Thomas, Winston
APPLICANT: Tsuchihashi, Zenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1
TITLE OF INVENTION: Sequences and Antibodies Thereo
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Flits, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 017957-000100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 246240 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..246240
OTHER INFORMATION: /note= "HLA-H.CONTIG"
US-08-724-394A-21
Query Match 2.8%; Score 37.2; DB 2; Length 246240;
Best Local Similarity 84.0%; Pred. No. 8.5;
Matches 42; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
OY 912 AACCTTTGGAGACGATTTGAGTAATAATAAACTGCTCCCTTA 961
DB 154422 AATTCACAGAGACGATTTGAGTAATAATAAACTGCTCCCTTA 154471
RESULT 7
US-08-724-394A-22
Sequence 22, Application US/08724394A
Patent No. 5872237
GENERAL INFORMATION:
APPLICANT: Feder, John N.
APPLICANT: Kironmal, Gregory S.
APPLICANT: Lauer, Peter M.
APPLICANT: Ruddy, David A.
APPLICANT: Thomas, Winston
APPLICANT: Tsuchihashi, Zenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1
TITLE OF INVENTION: Sequences and Antibodies Thereo
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Flits, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 017957-000100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 246240 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..246240
OTHER INFORMATION: /note= "HLA-H.CONTIG"
US-08-724-394A-22
Query Match 2.8%; Score 37.2; DB 2; Length 246240;
Best Local Similarity 84.0%; Pred. No. 8.5;
Matches 42; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
OY 912 AACCTTTGGAGACGATTTGAGTAATAATAAACTGCTCCCTTA 961
DB 154422 AATTCACAGAGACGATTTGAGTAATAATAAACTGCTCCCTTA 154471
RESULT 8
US-09-004-838-91/C
Sequence 91, Application US/09004838
Patent No. 6350933
GENERAL INFORMATION:
APPLICANT: Michelmore, Richard W.
APPLICANT: Shen, Kathy
APPLICANT: Meyers, Blake
TITLE OF INVENTION: Procedures and Materials for
TITLE OF INVENTION: Confering Pest Resistance in Plants
NUMBER OF SEQUENCES: 140
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/004,838
FILING DATE: 09-JAN-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/781,734
FILING DATE: 10-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Einhorn, Gregory P.
REGISTRATION NUMBER: 38,440
REFERENCE/DOCKET NUMBER: 023070-07881005
TELECOMMUNICATION INFORMATION:

[REDACTED]

1

;; TITLE OF INVENTION: Application to produce nematode resistant plants
;; FILE REFERENCE: U-012084-2
;; CURRENT APPLICATION NUMBER: US/09/234,827B
;; CURRENT FILING DATE: 1999-01-21
;; PRIOR APPLICATION NUMBER: US 60/072,142
;; PRIOR FILING DATE: 1998-01-22
;; NUMBER OF SEQ ID NOS: 32
;; SOFTWARE: Patentln Ver. 2.1
;; SEQ ID NO 1
;; LENGTH: 8302
;; TYPE: DNA
;; ORGANISM: Arabidopsis thaliana
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (4038)..(4256)
;; NAME/KEY: CDS
;; LOCATION: (4807)..(5604)
;; NAME/KEY: CDS
;; LOCATION: (6777)..(6827)
;; NAME/KEY: CDS
;; LOCATION: (6919)..(7935)
;; US-09-234-827B-1

Query Match 2.7%; Score 36.2; DB 4; Length 8302;
Best Local Similarity 51.2%; Pred. No. 4.5;
Matches 11; Conservative 0; Mismatches 103; Indels 3; Gaps 1;

QY 683 CTCGATTTCTTTGTTTCTTTTAAAGACAGAGCTGTCTCAAAAAGACATTTA 742
DB 805 CTTATGTTGGATTTTATTAAGACAGACCCGAAACATCAATACATTA 746
QY 743 TCATTAATACATCTATTTAGCCCTTATCTTATCGAAGGACGTTTCTTTT 802
DB 745 ATTTTAAATACATTTTATCAATTTTAATTTCTGTA--TTATTTTCAATTTTA 689
QY 803 AAACAGTTAGTACTGATGTCACAGACAAATTTTCTGATCAGTAGTCCCTGTCAAC 862
DB 688 AATAATTAATATTTAAATGTCTCCAAATAAATGATCGTGTTTATTTGGTTTACC 629
QY 863 AGTACAAATGTGTTTCTTAAGGCGGAGAAACA 899
DB 628 TGTAAAGACTCTGTTTATCAAAAAAGAAAAA 592

RESULT 12
US-09-424-283-6
Sequence 6, Application US/09424283
Patent No. 6437219
GENERAL INFORMATION:
APPLICANT: Grimes, et al.
TITLE OF INVENTION: Sucrose binding proteins
FILE REFERENCE: 4630-50206
CURRENT APPLICATION NUMBER: US/09/424,283
CURRENT FILING DATE: 1999-11-19
PRIOR APPLICATION NUMBER: PCT/US98/10465
PRIOR FILING DATE: 1998-05-21
PRIOR APPLICATION NUMBER: US 60/047,568
PRIOR FILING DATE: 1997-05-22
NUMBER OF SEQ ID NOS: 15
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 6
LENGTH: 3718
TYPE: DNA
ORGANISM: Glycine max
US-09-424-283-6

Query Match 2.7%; Score 36; DB 4; Length 3718;
Best Local Similarity 47.0%; Pred. No. 3.8;
Matches 11; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY 983 TTTACTGTCTATTATATCCCTAGTCCAAAGTAAATATTTCTGATATG 1042
DB 3257 TGTAGCACTATTGATTTTCTTCCACATTTATGAGTGAATCAGTTAGAGAA 3316

QY 1043 CTTTATACCAAGAGCCTTATCAGCAGCTTCCAGACACATATAGCAGAACATA 1102
DB 3317 ATATTAAAAAATAATTAATTAAGAGAGCAGATTAATTAAGATACGAACCTC 3376
QY 1103 CCACCAATACCAACATATAGCTTTTAATATCATGATAGCAGAGCATGATAT 1162
DB 3377 ACAATGATAGACCAATTAAGACATTTATTTCTTCAAAATTAAGAAAGCTTTT 3436
QY 1163 TGAGGCTTGATTTGGAAATACGACTCTACATTCATTTCAATCTTCAATACAT 1218
DB 3437 TTAACAATATATACATATATCATATATATTTATTTATTTATTTTAAACTT 3492

RESULT 13
US-09-009-913-1
Sequence 1, Application US/09009913
Patent No. 6087485
GENERAL INFORMATION:
APPLICANT: Arys Pharmaceuticals, Inc.
TITLE OF INVENTION: Asthma Related Genes
NUMBER OF SEQUENCES: 339
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bozicevic & Reed, LLP
STREET: 285 Hamilton Ave, Suite 200
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/009,913
FILING DATE: 21-JAN-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Sherwood, Pamela J
REGISTRATION NUMBER: 36,677
REFERENCE/DOCKET NUMBER: SEQ-4P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-327-3231
TELEFAX: 650-327-3231
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 72928 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-09-009-913-1

Query Match 2.7%; Score 35.8; DB 3; Length 72928;
Best Local Similarity 54.1%; Pred. No. 13;
Matches 73; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 708 AGACAGAGCTCTGTCTCAAAAAAGACATTTATCATTTAAATCTTTAGAGCC 767
DB 68237 AGAGTGAAGACTCTGTCAAAAAATTAATAAAATCACTGTATTTACTTTTG 68296
QY 768 CTAATTTCTTCTGAGAGGCACTGTTTCTTTTAAACGTTAGTACGATGCAACA 827
DB 68297 CACCAACATATATGATATATCAGACATTTATTTTAAAAAGTATTTGACATTTGCTTTTAA 68356
QY 828 GACAAATATTTCTGA 842
DB 68357 TATTAATTTTAA 68371

RESULT 14

US-09-134-001C-920/C

Sequence 920, Application US/09134001C

Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

FILE REFERENCE: GTC-007

CURRENT APPLICATION NUMBER: US/09/134,001C

CURRENT FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: US 60/055,779

NUMBER OF SEQ ID NOS: 5674

SEQ ID NO 920

LENGTH: 3780

TYPE: DNA

ORGANISM: Staphylococcus epidermidis

US-09-134-001C-920

Query Match

Best Local Similarity 50.6%; Pred. No. 4.9; DB 4; Length 3780;

Matches 86; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 273 TTTGAGAGTACTTTCTTTTGACGAAATTTTCATCTGCTGCGCATATTCCTCC 332

DB 3049 TTTCAACAGCAGCTTTCTTGAATATCTTTAGTCCATACCTGTTGATTTCTCAAT 2990

QY 333 CTTTATAGGAGCATGATTTCTTCTTTGTTGGAATGTCCTCATAGCATTTTCAG 392

DB 2989 CTTTATAGTACTTTTGTGATTAATTTGCGCTTTGATGAGATATATCAAAATGACAT 2930

QY 393 ATCTTTGATGTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 442

DB 2929 CAGCTACTTTCTGCTGCTGATCTATCTGCTGCTGCTGCTGCTGCTGCTGCTG 2880

RESULT 15

US-08-487-826B-13/C

Sequence 13, Application US/08487826B

Patent No. 5993827

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.

APPLICANT: Chitnis, Chetan

APPLICANT: Miller, Louis H.

APPLICANT: Peterson, David S.

APPLICANT: Su, Xin-zhaun

APPLICANT: Wellens, Thomas E.

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX

TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

NUMBER OF SEQUENCES: 45

CORRESPONDENCE ADDRESS:

ADDRESSEE: Knobe Martens Olson & Bear

CITY: Newport Beach

STATE: California

COUNTRY: US

ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08-487,826B

FILING DATE: 10-SEP-1993

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Israelson, Ned

REGISTRATION NUMBER: 29,655

REFERENCE/DOCKET NUMBER: NIH121.001CPI

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 235-8550

TELEFAX: (619) 235-0176

INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:

LENGTH: 19124 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

US-08-487-826B-13

Query Match

Best Local Similarity 45.0%; Pred. No. 8.8; DB 2; Length 19124;

Matches 134; Conservative 0; Mismatches 164; Indels 0; Gaps 0;

QY 903 TTTTAAAGTAACTTTTGGGAGACTGATTTGATATATATAAAGTCTGCTCCCTTAA 962

DB 6981 TTTTAAAGTAACTTTTGGGAGACTGATTTGATATATATAAAGTCTGCTCCCTTAA 9622

QY 963 GAAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1022

DB 6921 AAAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6862

QY 1023 TATCTATTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1082

DB 6861 TATCTATTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6802

QY 1083 CCAGTATACGACCAACATACCAACATACCAACATACCAACATACCAACATACCA 1142

DB 6801 AATACGAAATATCAAAATATCAAAATATCAAAATATCAAAATATCAAAATATCA 6742

QY 1143 TACAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1200

DB 6741 TACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6684

Search completed: June 27, 2003, 16:19:53
Job time: 97 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2003, 14:32:30 ; Search time 223 Seconds

(without alignments)
8860.093 Million cell updates/sec

Title: US-09-830-244B-2

Perfect score: 1331
Sequence: 1 ctatgtttttagaatcaag.....gccttttcaatgtacttg 1331

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1055720 seqs, 742224136 residues

Total number of hits satisfying chosen parameters: 2111440

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published_Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	104.6	7.9	438	US-09-924-340-5	Sequence 5, Appl1
2	104.6	7.9	438	US-09-992-600A-5	Sequence 5, Appl1
3	104.6	7.9	438	US-10-000-489-5	Sequence 5, Appl1
4	104.6	7.9	438	US-10-000-986-5	Sequence 5, Appl1
5	86.6	5.5	387	US-09-918-965-18537	Sequence 18537, A
6	74.8	5.5	433	US-09-983-965-5821	Sequence 5821, Ap
7	46.8	3.5	3104	US-10-037-596-32	Sequence 32, Appl
8	46.8	3.5	513509	US-09-754-853A-4	Sequence 4, Appl1
9	43.2	3.2	65359	US-09-804-472-3	Sequence 3, Appl1
10	43.2	3.2	6055	US-10-091-504-1746	Sequence 1746, Ap
11	42.2	3.2	6055	US-09-764-869-1746	Sequence 1746, Ap
12	42.2	3.2	7809	US-10-091-504-1744	Sequence 1744, Ap
13	42.2	3.2	7809	US-10-091-504-1745	Sequence 1745, Ap
14	42.2	3.2	7809	US-09-764-869-1744	Sequence 1744, Ap
15	42.2	3.2	7809	US-09-764-869-1745	Sequence 1745, Ap
16	41.8	3.1	7934	US-10-239-676-154	Sequence 154, App
17	40.6	3.1	556	US-10-092-154-1685	Sequence 1685, Ap
18	40.6	3.1	556	US-09-764-847-1685	Sequence 1685, Ap
19	40.6	3.1	175561	US-10-017-721-3	Sequence 3, Appl1

C 20	40.4	3.0	17849	9	US-10-092-154-1315	Sequence 1315, Ap
C 21	40.4	3.0	17849	9	US-09-764-891-10139	Sequence 10139, A
C 22	40.4	3.0	17849	10	US-09-764-847-1315	Sequence 1315, Ap
C 23	40.4	3.0	17862	9	US-10-092-154-1313	Sequence 1313, Ap
C 24	40.4	3.0	17862	9	US-09-764-891-10138	Sequence 10138, A
C 25	40.4	3.0	17862	10	US-09-764-847-1313	Sequence 1313, Ap
C 26	40.4	3.0	11047	9	US-10-239-676-188	Sequence 188, App
C 27	39.6	3.0	151	10	US-09-864-761-22646	Sequence 22646, A
C 28	39.6	3.0	492	10	US-09-864-761-5885	Sequence 5885, Ap
C 29	39.6	3.0	2000	9	US-09-938-842A-4305	Sequence 4305, Ap
C 30	39.6	3.0	14147	9	US-10-172-086-52	Sequence 52, Appl
C 31	39.6	3.0	44848	9	US-09-968-113-42	Sequence 42, Appl
C 32	39.6	3.0	44848	10	US-09-776-874A-42	Sequence 42, Appl
C 33	39.4	3.0	1691139	9	US-10-067-514-1	Sequence 1, Appl1
C 34	39.2	2.9	5867	9	US-10-239-676-37	Sequence 37, Appl
C 35	39.2	2.9	14147	9	US-10-172-086-52	Sequence 52, Appl
C 36	39.2	2.9	378361	9	US-09-901-136-3	Sequence 3, Appl
C 37	38.6	2.9	5236	10	US-09-925-297-363	Sequence 363, App
C 38	38.6	2.9	6265	10	US-09-129-112-3	Sequence 61, Appl
C 39	38.6	2.9	10758	12	US-10-044-090-61	Sequence 3, Appl1
C 40	38.4	2.9	472	9	US-09-918-995-3775	Sequence 3775, Ap
C 41	38.2	2.9	2000	9	US-09-938-842A-3631	Sequence 3631, Ap
C 42	38.2	2.9	2286	9	US-10-228-796-3	Sequence 3, Appl1
C 43	38.2	2.9	2286	10	US-09-191-687B-3	Sequence 97, Appl
C 44	38.2	2.9	7906	9	US-10-239-676-97	Sequence 133, App
C 45	38.2	2.9	588	9	US-09-871-161-133	

ALIGNMENTS

RESULT 1
US-09-924-340-5
Sequence 5, Application US/09924340
Publication No. US20030027248A1
GENERAL INFORMATION:
APPLICANT: Bejanin, Stephanie
TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
FILE REFERENCE: 91, US2, REG
CURRENT APPLICATION NUMBER: US/09/924, 340
CURRENT FILING DATE: 2001-08-06
PRIORITY APPLICATION NUMBER: US 60/305, 456
PRIORITY FILING DATE: 2001-07-13
PRIORITY APPLICATION NUMBER: US 60/302, 277
PRIORITY FILING DATE: 2001-06-29
PRIORITY APPLICATION NUMBER: US 60/298, 698
PRIORITY FILING DATE: 2001-06-15
PRIORITY APPLICATION NUMBER: US 60/293, 574
PRIORITY FILING DATE: 2001-05-25
NUMBER OF SEQ ID NOS: 112
SOFTWARE: Jpatent
SEQ ID NO 5
LENGTH: 438
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: 5'UTR
LOCATION: 1..83
NAME/KEY: CDS
LOCATION: 84..317
NAME/KEY: 3'UTR
LOCATION: 318..438
NAME/KEY: polyA_signal
LOCATION: 397..402
NAME/KEY: polyA_site
LOCATION: 423..438
US-09-924-340-5

Query Match 7.9% Score 104.6; DB 9; Length 438;
Best Local Similarity 88.3%; Pred. No. 7.6e-17;
Matches 113; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

PRIOR APPLICATION NUMBER: US 09/924,340
PRIOR FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: PCT/IB01/01715
PRIOR FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: US 60/305,456
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/302,277
PRIOR FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: US 60/298,698
PRIOR FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: US 60/293,574
PRIOR FILING DATE: 2001-05-25
NUMBER OF SEQ ID NOS: 112
SOFTWARE: Jpatent
SEQ ID NO 5
LENGTH: 438
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: 5'UTR
LOCATION: 1..83
NAME/KEY: CDS
LOCATION: 84..317
NAME/KEY: 3'UTR
LOCATION: 318..438
NAME/KEY: polyA_signal
LOCATION: 397..402
NAME/KEY: polyA_site
LOCATION: 423..438
US-10-000-986-5

Query Match
Best Local Similarity 88.3%; Pred. No. 7.6e-17;
Matches 113; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1204 CACTTTCATACATATACACTACTACCACTTTTGTGATCATCTAGACCAATGCG 1263
DB 192 CACTATTCATACATATACACTACTACCACTTTTGTGATCATCTAGACCAATGCG 251
QY 1264 AATGTAAACCCATATATATCTGATGATCTGCTGATCTGCTGCTTTTCAAT 1323
DB 252 AATGAAAAACATATATATCTGATGATCTGCTGATCTGCTGCTTTTCAAT 311
QY 1324 GTCACTTG 1331
DB 312 GTCACTTG 319

RESULT 5
US-09-918-995-18537/c
Sequence 18537, Application US/09918995
Publication No. US20030073623A1
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
FILE REFERENCE: 20411-756
CURRENT APPLICATION NUMBER: US/09/918,995
PRIOR FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US/09/235,076
PRIOR FILING DATE: 1999-01-20
NUMBER OF SEQ ID NOS: 38054
SOFTWARE: fastseq for windows Version 3.0
SEQ ID NO 18537
LENGTH: 387
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(387)
OTHER INFORMATION: n = A,T,C or G
US-09-918-995-18537

Query Match
Best Local Similarity 6.5%; Score 86.6; DB 9; Length 387;
Matches 122; Conservative 0; Mismatches 34; Indels 5; Gaps 1;

QY 812 AGTCTGATGTCACAGACAAATATTTCTGATCAGATAGTCCCTGTCACAGTACGAA 871
DB 345 AGAAACATGAAATTTTCCAAATATTTCCGATCAGAGATC-----ACAAAGCAGCAGAA 291
QY 872 TGTGCTTCATTAAGTGGGAAAGAAACAGCATTTTAAAGTACTTTTGGGAGACTGAT 931
DB 290 TGTGCTTCATCAGTGGGAAAGAAAGCAGCATTTTAAATGACTTTTGGGAGACTGAT 231
QY 932 TGAGTAAATTAATAAACTCTGCTCCTTAAAGAAAAAAA 972
DB 230 TGAGTAAATTAATAAACTCTGCTCCTTAAATTAATAA 190

RESULT 6

US-09-983-965-5821
Sequence 5821, Application US/09983965
Patent No. US20020137160A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Iao, Nengbing
APPLICANT: Byatt, John C.
APPLICANT: Mathalagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 37-21(10297)/C
CURRENT APPLICATION NUMBER: US/09/983,965
PRIOR FILING DATE: 2001-10-26
PRIOR APPLICATION NUMBER: US 09/465,231
PRIOR FILING DATE: 1999-12-15
PRIOR APPLICATION NUMBER: US 60/113,678
PRIOR FILING DATE: 1998-12-17
NUMBER OF SEQ ID NOS: 5912
SEQ ID NO 5821
LENGTH: 433
TYPE: DNA
ORGANISM: Bos taurus
FEATURE:
OTHER INFORMATION: Clone ID: 60-LIB34-039-Q1-E1-G8
US-09-983-965-5821

Query Match
Best Local Similarity 5.6%; Score 74.8; DB 10; Length 433;
Matches 142; Conservative 0; Mismatches 62; Indels 10; Gaps 2;

QY 482 TTGATTTGCTGACCTCTACCAATAGCCTTTTGAATGACGAAAGTGAAGAGAA 541
DB 4 TGGGTTTCCTGCTACTCTCTCAAGAGCCTTTGAATGACTGGA-----ATGTCACAGA 57
QY 542 GAGCATGCTGCAAGAAAGATAGCTAATATTTTGGTACTTATCTGAATCCAGA 601
DB 58 GAAAGAACTGTCAGAAAGAGTATTTTGGAGAGCTTGAATCCAGA 117
QY 602 TGTGCTTCCTCCCTGAGGTGTTTCTTACGATCCTCATTAATCCCTCGGAG 661
DB 118 TGTGCTTCCTCCCTGAGGTGTTTCTTACGATCCTCATTAATCCCTCGGAG 661
QY 662 CACAGAGAGTTAGTAGACACTCTCCATTTCTTTG 695
DB 174 GATAGCCCGACGAGTAGAAGTGTATTTCTTG 207

RESULT 7

US-10-037-598-32
Sequence 32, Application US/10037598
Patent No. US20020157143A1
GENERAL INFORMATION:
APPLICANT: Monsanto Co
APPLICANT: Conclido, Vergel
APPLICANT: Delaney, Xavier

CURRENT FILING DATE: 2002-03-07
NUMBER OF SEQ ID NOS: 2442
Prior Application removed - See File Wrapper or Palm
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1746
LENGTH: 6055
TYPE: DNA
ORGANISM: Homo sapiens
US-10-091-504-1746

Query Match
Best Local Similarity 45.8%; Score 42.2; DB 9; Length 6055;
Pred. No. 3.2;
Matches 143; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

QY 937 AATATAAACTCTGCTCCCTTAAGAAAAAACCCCTTCACCTTACTGTGCATT 996
DB 835 AAAGAAAGGATACCTTTCTTAAATGAATGTAACCCGTAAGAAATGTTTAA 894
QY 997 TATATCCCTTATGTCCTCAAGTTAATATCTTATTCGGTATTCCTTTATACCAAG 1056
DB 895 AAGACTGCCAGACTCTTTAAGGCTTTAAGCTTATGTTATCTCTTTTATCTACT 954
QY 1057 AGCCTTATCAGCAGCTTCAGAACACACACTATACGCAACATACCAACCAATACC 1116
DB 955 ATAGAAATAGAAATTAAGTGAAGAAAGTAATTAATTAATTTTGTACATTTT 1014
QY 1117 AACAAATAGCTTTAATATCATCAGTAAGTGCAGCATGATTTAGGCTTGATTTG 1176
DB 1015 AAATTAATATGATTTCTTATATGTAATTAATTAATTAATTAATTTTAAAT 1074
QY 1177 CAATATGACTCTTACATTCATCTTCTTACATCAGTATTCACATACCACTT 1236
DB 1075 GTAAACAACTTTAAGTACAGCCTTATATTTAATTTCTGTAGGCTAGAAATCAGATT 1134
QY 1237 TTGTGNAGTCA 1248
DB 1135 AGTGCCATAGCA 1146

RESULT 11

US-09-764-869-1746
Sequence 1746, Application US/09764869
Patent No. US20020061521A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC007
CURRENT APPLICATION NUMBER: US/09/764,869
CURRENT FILING DATE: 2001-01-17
Prior Application data removed - refer to PALM or file wrapper
NUMBER OF SEQ ID NOS: 2442
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1746
LENGTH: 6055
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-869-1746

Query Match
Best Local Similarity 45.8%; Score 42.2; DB 10; Length 6055;
Pred. No. 3.2;
Matches 143; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

QY 937 AATATAAACTCTGCTCCCTTAAGAAAAAACCCCTTCACCTTACTGTGCATT 996
DB 835 AAAGAAAGGATACCTTTCTTAAATGAATGTAACCCGTAAGAAATGTTTAA 894
QY 997 TATATCCCTTATGTCCTCAAGTTAATATCTTATTCGGTATTCCTTTATACCAAG 1056
DB 895 AAGACTGCCAGACTCTTTAAGGCTTTAAGCTTATGTTATCTCTTTTATCTACT 954
QY 1057 AGCCTTATCAGCAGCTTCAGAACACCACTATACGCAACATACCAACCAATACC 1116
DB 955 ATAGAAATAGAAATTAAGTGAAGAAAGTAATTAATTAATTTTGTACATTTT 1014

QY 1117 AACATATAGCTTTAATATCATCAGTAAGTGCAGCATGATTTAGGCTTGATTTG 1176
DB 1015 AAATTAATATGATTTCTTATATGTAATTAATTAATTAATTAATTTTAAAT 1074
QY 1177 CAATATGACTCTTACATTCATCTTCTTACATCAGTATTCACATACCACTT 1236
DB 1075 GTAAACAACTTTAAGTACAGCCTTATATTTAATTTCTGTAGGCTAGAAATCAGATT 1134
QY 1237 TTGTGNAGTCA 1248
DB 1135 AGTGCCATAGCA 1146

RESULT 12

US-10-091-504-1744
Sequence 1744, Application US/10091504
Publication No. US20030059908A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC007C1
CURRENT APPLICATION NUMBER: US/10/091,504
CURRENT FILING DATE: 2002-03-07
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 2442
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1744
LENGTH: 7809
TYPE: DNA
ORGANISM: Homo sapiens
US-10-091-504-1744

Query Match
Best Local Similarity 45.8%; Score 42.2; DB 9; Length 7809;
Pred. No. 3.7;
Matches 143; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

QY 937 AATATAAACTCTGCTCCCTTAAGAAAAAACCCCTTCACCTTACTGTGCATT 996
DB 2277 AAAGAAAGGATACCTTTCTTAAATGAATGTAACCCGTAAGAAATGTTTAA 2336
QY 997 TATATCCCTTATGTCCTCAAGTTAATATCTTATTCGGTATTCCTTTATACCAAG 1056
DB 2337 AAGACTGCCAGACTCTTTAAGGCTTTAAGCTTATGTTATCTCTTTTATCTACT 2396
QY 1057 AGCCTTATCAGCAGCTTCAGAACACCACTATACGCAACATACCAACCAATACC 1116
DB 2397 ATAGAAATAGTAATTAAGTGAAGAAAGTAATTAATTAATTTTGTACATTTT 2456
QY 1117 AACAAATAGCTTTAATATCATCAGTAAGTGCAGCATGATTTAGGCTTGATTTG 1176
DB 2457 AAATTAATATGATTTCTTATATGTAATTAATTAATTAATTAATTTTAAAT 2516
QY 1177 CAATATGACTCTTACATTCATCTTCTTACATCAGTATTCACATACCACTT 1236
DB 2517 GTAAACAACTTTAAGTACAGCCTTATATTTAATTTCTGTAGGCTAGAAATCAGATT 2576
QY 1237 TTGTGNAGTCA 1248
DB 2577 AGTGCCATAGCA 2588

RESULT 13

US-10-091-504-1745
Sequence 1745, Application US/10091504
Publication No. US20030059908A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC007C1
CURRENT APPLICATION NUMBER: US/10/091,504
CURRENT FILING DATE: 2002-03-07
NUMBER OF SEQ ID NOS: 2442

;; Prior Application removed - See file Wrapper or Palm
;; SOFTWARE: Patentin Ver. 2.0
;; SEQ ID NO 1745
;; LENGTH: 7809
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-10-091-504-1745

Query Match
Best Local Similarity 45.8%; Score 42.2; DB 9; Length 7809;
Matches 143; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

QY 937 AATAATAAACTGCTGCTCCCTTAAGAAAAAACCCCTCCACCTTTACTGTGTCATT 996
DB 2276 AAGAAAGAGATCTTTCTTATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 2335
QY 997 TATATCCCTTAGTTCCTCAAGTTAATATCTTATTTCTGATATTTGCTTTTAAACCAAG 1056
DB 2336 AAGACTGCCAGACTTTTAAAGGCTTTAACTATTTATTTATCTTCTTTTATTTCTAACT 2395
QY 1057 AGCCTTATCAGCCAGTTCCAGACACCACTATTCGACACCACTACCAACCAATACC 1116
DB 2396 ATAGAAATATGATATTTATCTGTAAGAAAGATGATTAATAATTTATTTTGTACATTTT 2455
QY 1117 AACATATPAGCTTTTAAATATCATCATGTAAGTACGAGACATGATTTAGGCTTGAATGG 1176
DB 2456 AATAATATATGATTTCTTATATGTAATATAAATGTAATGTAATGTAATGTAATGTAAT 2515
QY 1177 CAATATAGACTTCTACATCATATTTCTATCTTATCATATTCATACACTACTACCACTT 1236
DB 2516 GTAAACACACTTTTAACTAGTACAGCCTTATTTTAAATTTCTGTAGGCTAGAAATCAGAAAT 2575
QY 1237 TTGTNAGATCA 1248
DB 2576 AGTCCATATGCA 2587

RESULT 14

US-09-764-869-1744
;; Sequence 1744, Application US/09764869
;; Patent No. US20020061521A1
;; GENERAL INFORMATION:
;; APPLICANT: Rosen et al.
;; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;; FILE REFERENCE: PC007
;; CURRENT APPLICATION NUMBER: US/09/764,869
;; CURRENT FILING DATE: 2001-01-17
;; Prior application data removed - refer to PALM or file wrapper
;; NUMBER OF SEQ ID NOS: 2442
;; SOFTWARE: Patentin Ver. 2.0
;; SEQ ID NO 1744
;; LENGTH: 7809
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-09-764-869-1744

Query Match
Best Local Similarity 45.8%; Score 42.2; DB 10; Length 7809;
Matches 143; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

QY 937 AATAATAAACTGCTGCTCCCTTAAGAAAAAACCCCTCCACCTTTACTGTGTCATT 996
DB 2277 AAGAAAGAGATCTTTCTTATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 2336
QY 997 TATATCCCTTAGTTCCTCAAGTTAATATCTTATTTCTGATATTTGCTTTTAAACCAAG 1056
DB 2337 AAGACTGCCAGACTTTTAAAGGCTTTAACTATTTATTTATCTTCTTTTATTTCTAACT 2396
QY 1057 AGCCTTATCAGCCAGTTCCAGACACCACTATTCGACACCACTACCAACCAATACC 1116
DB 2397 ATAGAAATATGATTTCTTATGTAAGAAAGATGATTAATAATTTATTTTGTACATTTT 2456
QY 1117 AACATATPAGCTTTTAAATATCATCATGTAAGTACGAGACATGATTTATTTGAGCTTGAATGG 1176

DB 2457 AATAATAATGATTTTCTTATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 2516
QY 1177 CAATATAGACTTCTACATCATATTTCTATCTTTATACATATCAGCTACTACCACTT 1236
DB 2517 GTAAACACACTTTTAACTAGTACAGCCTTATTTTAAATTTCTGTAGGCTAGAAATCAGAAAT 2576
QY 1237 TTGTNAGATCA 1248
DB 2577 AGTCCATATGCA 2587

RESULT 15

US-09-764-869-1745
;; Sequence 1745, Application US/09764869
;; Patent No. US20020061521A1
;; GENERAL INFORMATION:
;; APPLICANT: Rosen et al.
;; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;; FILE REFERENCE: PC007
;; CURRENT APPLICATION NUMBER: US/09/764,869
;; CURRENT FILING DATE: 2001-01-17
;; Prior application data removed - refer to PALM or file wrapper
;; NUMBER OF SEQ ID NOS: 2442
;; SOFTWARE: Patentin Ver. 2.0
;; SEQ ID NO 1745
;; LENGTH: 7809
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-09-764-869-1745

Query Match
Best Local Similarity 45.8%; Score 42.2; DB 10; Length 7809;
Matches 143; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

QY 937 AATAATAAACTGCTGCTCCCTTAAGAAAAAACCCCTCCACCTTTACTGTGTCATT 996
DB 2276 AAGAAAGAGATCTTTCTTATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 2335
QY 997 TATATCCCTTAGTTCCTCAAGTTAATATCTTATTTCTGATATTTGCTTTTAAACCAAG 1056
DB 2336 AAGACTGCCAGACTTTTAAAGGCTTTAACTATTTATTTATCTTCTTTTATTTCTAACT 2395
QY 1057 AGCCTTATCAGCCAGTTCCAGACACCACTATTCGACACCACTACCAACCAATACC 1116
DB 2396 ATAGAAATATGATTTCTTATGTAAGAAAGATGATTAATAATTTATTTTGTACATTTT 2455
QY 1117 AACATATPAGCTTTTAAATATCATCATGTAAGTACGAGACATGATTTAGGCTTGAATGG 1176
DB 2456 AATAATATGATTTTCTTATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 2515
QY 1177 CAATATAGACTTCTACATCATATTTCTATCTTTATACATATCAGCTACTACCACTT 1236
DB 2516 GTAAACACACTTTTAACTAGTACAGCCTTATTTTAAATTTCTGTAGGCTAGAAATCAGAAAT 2575
QY 1237 TTGTNAGATCA 1248
DB 2576 AGTCCATATGCA 2587

Search completed: June 27, 2003, 16:23:52
Job time: 229 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2003, 12:58:01 ; Search time 1825 seconds
(without alignments)

11811.613 Million cell updates/sec

Title: US-09-830-244b-2

Perfect score: 1331

Sequence: 1 catgttttagatcaaga.....gcctttccaatgtcactgt 1331

Scoring table:

IDENTITY_NIC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 809774376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estmu:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_estl:*
9: gb_estl:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	490	36.8	492	9	AI224097
C 2	480	36.1	518	9	AI694371
C 3	478.4	35.9	523	10	AM269813
C 4	405	30.4	888	12	BG121534
C 5	399	30.0	427	12	BF831772
C 6	398.4	29.9	417	12	BE767736

C 7	380.4	28.6	394	12	BG011356
C 8	369.4	27.8	403	12	BF831960
C 9	353	26.5	408	12	BG011354
C 10	352.8	26.5	392	12	BE297971
C 11	345.6	26.0	402	12	BG011351
C 12	344.8	25.9	395	12	BF96088
C 13	335.2	25.2	951	12	BF168510
C 14	333	25.0	396	12	BF832034
C 15	330.4	24.8	376	9	AA488304
C 16	322	24.2	386	14	BQ359015
C 17	309.2	23.2	352	14	BQ359019
C 18	293.2	22.0	324	12	BF986192
C 19	280.8	21.1	417	14	BQ359018
C 20	266.6	20.0	295	10	AM135014
C 21	262.4	19.7	279	12	BF19548
C 22	262.4	19.0	365	10	BE089961
C 23	252.6	18.8	517	12	BG188342
C 24	249.8	18.7	470	12	BG200817
C 25	248.8	18.5	404	9	AA376713
C 26	246.2	18.4	405	9	AA376719
C 27	245.2	18.2	519	10	AM951480
C 28	242.4	18.0	250	9	AI630950
C 29	239	17.9	410	14	T29650
C 30	238.4	17.7	335	9	AA376681
C 31	236	17.2	375	9	AA376714
C 32	228.6	16.7	257	12	BF832847
C 33	222	16.7	457	9	AA376682
C 34	211	15.9	457	9	AA376682
C 35	203.4	15.3	306	10	BE163886
C 36	197.8	14.9	426	9	AA376680
C 37	181.4	13.6	302	9	AA376694
C 38	179.2	13.5	341	9	AA376695
C 39	175.6	13.2	329	14	BQ309679
C 40	169.6	12.7	321	12	BG205925
C 41	163	12.2	258	12	BF834202
C 42	161.4	12.1	314	10	AM409123
C 43	161	12.1	209	12	BF835534
C 44	155.6	11.7	173	12	BF088945
C 45	155	11.6	172	12	BF944749

ALIGNMENTS

RESULT 1
AI224097 492 bp mRNA linear EST 28-JAN-1999
LOCUS g135a03.x1 Soares_NHMPu_S1 Homo sapiens cDNA clone IMAGE:1858444
DEFINITION 3', mRNA sequence.

ACCESSION AI224097
VERSION AI224097.1 GI:3806810
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 492)
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
NCI-CCAP <http://www.ncbi.nlm.nih.gov/nciccap>.

AI224097
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-f@mail.nih.gov
This clone is available royalty-free through LNLN; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 876 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 456.

FEATURES

source
1. 492
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1858444"

Query Match	36.8%	Score 490;	DB 9;	Length 492;
Best Local Similarity	99.6%	Pred. No. 8.1e-105;		
Matches 490;	Conservative	0;	Mismatches 2;	Indels 0;
				Gaps 0

TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP).									
JOURNAL	Tumor Gene Index									
COMMENT	Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: M. Bento Soares, Ph.D. CDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/FLNLT at: www.bio.lnhi.gov/bbrp/image/image.html Insert Length: 710 Std Error: 0.00 Seq primer: -40up from Gibco High quality sequence stop: 454. Location/Qualifiers 1. 518 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:2338158" /clone_lib="NCI-CGAP-Lu24" /tisue_type="carcinoid" /lab_host="DH10B" /note="Organ: lung; Vector: pRTD3-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI-CGAP Lu5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonoids 141920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Boraldo."									
BASE COUNT	171 a	107 c	83 g	157 t						
ORIGIN										
Query Match	36.1%	Score 480;	DB 9;	Length 518;						
Best Local Similarity	100.0%;	Pred. No. 1.8e-102;								
Matches 480;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;						
OY	1	CTATGTTTATAGATCAAAAGTGAACCGGGTACGTCTCATGTACCAACGTGAATTT	60							
Db	489	CTATGTTTATAGATCAAAAGTGAACCGGGTACGTCTCATGTACCAACGTGAATTT	60							
OY	61	ACAGTGTTTACAATGCTCTGGAATTTTGGACCTGCCATAGGGAATGTTAAGTTACTTGGC	120							
Db	429	ACAGTGTTTACAATGCTCTGGAATTTTGGACCTGCCATAGGGAATGTTAAGTTACTTGGC	120							
OY	121	TGGATTTATCAGACTTGTGAGTAAACACTGTGAAGTTTATAGCAGATGAGGGGGAATATTTG	180							
Db	369	TGGATTTATCAGACTTGTGAGTAAACACTGTGAAGTTTATAGCAGATGAGGGGGAATATTTG	180							
OY	181	AGGCCCCCTAAGGCTAAACAATAATACAGATGTGAGATGTGGGCAATCTGGCTCCCA	240							
Db	309	AGGCCCCCTAAGGCTAAACAATAATACAGATGTGAGATGTGGGCAATCTGGCTCCCA	240							
OY	241	GGCCTAATTTGGGACAGTTTTTCCGTGATTCCTTTGAGAGATCTTCTTTGACAGAA	300							
Db	249	GGCCTAATTTGGGACAGTTTTTCCGTGATTCCTTTGAGAGATCTTCTTTGACAGAA	300							
OY	301	TTTTCAATTCGCTGGCATTCGATATATTCCTCTTATAGAGCCATTGGATTTCTTCC	360							
Db	189	TTTTCAATTCGCTGGCATTCGATATATTCCTCTTATAGAGCCATTGGATTTCTTCC	360							
OY	361	TTTTTGGGGAATGTCCCATTTAGCATTTTTCAGATCTTTGATGTGACATTAATGCCATTAT	420							
Db	129	TTTTTGGGGAATGTCCCATTTAGCATTTTTCAGATCTTTGATGTGACATTAATGCCATTAT	420							
OY	421	TGCTATGCCGTTATTGGTGAATACGCAATAGTTAAATAACTGTTTACAGTAATCTACA	480							
Db	69	TGCTATGCCGTTATTGGTGAATACGCAATAGTTAAATAACTGTTTACAGTAATCTACA	480							

```

RESULT 3
AM269813/c 523 bp mRNA linear EST 03-JAN-2000
LOCUS xv45f02.x1 Soares_NFL_T_GBC_SI Homo sapiens cDNA clone
DEFINITION AM269813
ACCESSION AM269813
VERSION AM269813.1 GI:6656843
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 523)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: -40up from gibco
High quality sequence stop: 462.
FEATURES
SOURCE
1..523
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2816091"
/clone.lib="Soares_NFL_T_GBC_SI"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site:1: Not I; Site:2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL9W, testis NMH, and B-cell
NCI CGAP GCBI) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 177 a 109 c 85 g 132 t
ORIGIN
Query Match 35.9%; Score 478.4; DB 10; Length 523;
Best Local Similarity 99.8%; Pred. No. 4.3e-102;
Matches 479; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CTATGTTTGAATCAAGATGACCGGTAAGCTGTCATGTACCAACGTAATTT 60
DB 481 CTATGTTTGAATCAAGATGACCGGTAAGCTGTCATGTACCAACGTAATTT 422
QY 61 ACAGTGTTCACAAATGCTGGAATTTGCACCTGCCATAGGGAATGTTAAGCTTGGC 120
DB 421 ACAGTGTTCACAAATGCTGGAATTTGCACCTGCCATAGGGAATGTTAAGCTTGGC 362
QY 121 TGAATTTATCAGACTGTGAGTAACAAGTTGAAGTTAGCAGATAGGGGGGAATTG 180
DB 361 TGAATTTATCAGACTGTGAGTAACAAGTTGAAGTTAGCAGATAGGGGGGAATTG 302
QY 181 AGGCCCCCTAAGGCTAAACAAATATATCATGATAGTGGCTAATGTGCTCCCA 240
DB 301 AGGCCCCCTAAGGCTAAACAAATATATCATGATAGTGGCTAATGTGCTCCCA 242
QY 241 GGCCTAATTTGGGAACAGTTTTCGATGCTTGTAGAGTACTTTCTTTACAGAAA 300
DB 241 GGCCTAATTTGGGAACAGTTTTCGATGCTTGTAGAGTACTTTCTTTACAGAAA 182
QY 301 TTTTCATTTCTGCTTGCATTTCTATATTTCTCCCTTTATAGAGCCATTGGATTCTTTC 360
DB 181 TTTTCATTTCTGCTTGCATTTCTATATTTCTCCCTTTATAGAGCCATTGGATTCTTTC 122

```

```

QY 361 TTTTGGGAATATGTCATTTAGCATTTTCAGATCTTTGATGTCAGTATGCAATAT 420
DB 121 TTTTGGGAATATGTCATTTAGCATTTTCAGATCTTTGATGTCAGTATGCAATAT 62
QY 421 TGGTAATGCCGTTATTTGTAATACAGCATAGTAAATTAAGTGTACAGTAATCTACA 480
DB 61 TGGTAATGCCGTTATTTGTAATACAGCATAGTAAATTAAGTGTACAGTAATCTACA 2
RESULT 4
BG121534 888 bp mRNA linear EST 30-JAN-2001
LOCUS 602352872P1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4451201 5'
DEFINITION BG121534
ACCESSION BG121534
VERSION BG121534.1 GI:12615043
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 888)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LLM10237 row: p column: 18
High quality sequence start: 2
High quality sequence stop: 680.
FEATURES
SOURCE
1..888
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4451201"
/clone.lib="NIH_MGC_90"
/tissue-type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: PCMV-SpOrt6; Site:1: NotI;
Site:2: SalI; cloned unidirectionally; oligo-dT primed.
Average insert size 1.7 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH-MGC Library."
BASE COUNT 240 a 161 c 198 g 289 t
ORIGIN
Query Match 30.4%; Score 405; DB 12; Length 888;
Best Local Similarity 89.8%; Pred. No. 7.9e-85;
Matches 468; Conservative 0; Mismatches 50; Indels 3; Gaps 3;
QY 1 CTATGTTTGAATCAAGATGACCGGTAAGCTGTCATGTACCAACGTAATTT 60
DB 284 CTATGTTTGAATCAAGATGACCGGTAAGCTGTCATGTACCAACGTAATTT 343
QY 61 ACAGTGTTCACAAATGCTGGAATTTGCACCTGCCATAGGGAATGTTAAGCTTGGC 120
DB 344 ACAGTGTTCACAAATGCTGGAATTTGCACCTGCCATAGGGAATGTTAAGCTTGGC 403
QY 121 TGAATTTATCAGACTGTGAGTAACAAGTTGAAGTTAGCAGATAGGGGGGAATTG 180
DB 404 TGAATTTATCAGACTGTGAGTAACAAGTTGAAGTTAGCAGATAGGGGGGAATTG 463
QY 181 AGGCCCCCTAAGGCTAAACAA-ATATCATGTAATGAGATAGGCTAATGTGCTCCG 239
DB 464 AGGCCCCCTAAGGCTAAACAAACATATATCATGTAATGAGATAGGCTAATGTGCTCCG 523

```


QY 240 AGCCCTAATTTGGGACAGTCTTTCTGATGCTTGGAGAGTACTTCTTTGACAGAA 299
 Db 524 AGG-CTAATTTGGGACAGTCTTTCTGATGCTTGGAGAGTACTTCTTTGACAGAA 582
 QY 300 ATTTTCAATTCCTGCTTCCCATTTGCTATATTCCTTTATAGAGGACCATTTGCTTTC 359
 Db 583 ATTTTCAATTCCTGCTTCCCATTTGCTATATTCCTTTATAGAGGACCATTTGCTTTC 641
 QY 360 CTTTGTGGGAATGTCCTATAGCATTTTTCAGATCTTTTGTGACACTAATGCCATTA 419
 Db 642 CTTTGTGGGAATGTCCTATAGCATTTTTCAGATCTTTTGTGACACTAATGCCATTA 701
 QY 420 TTGCTAATGCTTATTTGCTGATACAGCATAGTAAATTAATCTTACAGTAATCTAC 479
 Db 702 ATGGTAATGCTTATTTGCTGATACAGCATAGTAAATTAATCTTACAGTAATCTAC 761
 QY 480 ACTTGATTTGCTGACCTCTTACAGCATAGTAAATTAATCTTACAGTAATCTAC 520
 Db 762 ACTTGATTTGCTGACCTCTTACAGCATAGTAAATTAATCTTACAGTAATCTAC 802
 RESULT 5
 LOCUS BF831772/c 427 bp mRNA linear EST 13-JAN-2001
 DEFINITION PM3-HT0909-151000-009-e12 HT0909 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BF831772
 VERSION BF831772.1 GI:12179873
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 427)
 AUTHORS Dias Neto,E., Garcia Correa,R., Veijovskij-Almeida,S., Bionesi,M.R.,
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
 Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
 'M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
 Simpson,A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?cl=PM3&t2=PM3-HT0909-
 151000-009-e12&t3=2000-10-15&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 18
 High quality sequence stop: 427.
 Location/Qualifiers
 1..427
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_id="HT0909"
 /dev_stage="Adult"
 /note="Organ: head,neck; Vector: puc18; Site:1: Sma1;
 Site:2: Sma1; A mini-library was made by cloning products
 derived from ORESSES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the puc 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."
 146 a 95 c 72 g 113 t 1 others
 BASE COUNT
 ORIGIN

Query Match 30.0%; Score 399; DB 12; Length 427;
 Best Local Similarity 99.8%; Pred. No. 2,1e-83;
 Matches 399; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 42 TGTACCAACGTAATTTTACAGTGTGTTTACAAATGCTGTGAAATTTTGCACCTCCATAGGG 101
 Db 427 TGTACCAACGTAATTTTACAGTGTGTTTACAAATGCTGTGAAATTTTGCACCTCCATAGGG 368
 QY 102 AATGTTAAGGTTACTTGGCGGTAATTTATCAGACTGTGTAGTAACAACTTGAAGTTTNG 161
 Db 367 AATGTTAAGGTTACTTGGCGGTAATTTATCAGACTGTGTAGTAACAACTTGAAGTTTNG 308
 QY 162 CAGATGAGGGGGAATTTATGAGCCCCCTAAGCCTTAACAAATTAATACAGATGCTGATG 221
 Db 307 CAGATGAGGGGGAATTTATGAGCCCCCTAAGCCTTAACAAATTAATACAGATGCTGATG 248
 QY 222 TGCTATATGCTGCTCCCGAGCCCTAATTTGGAGACAGTTTTCGTGATTCCTTTGAGAG 281
 Db 247 TGCTATATGCTGCTCCCGAGCCCTAATTTGGAGACAGTTTTCGTGATTCCTTTGAGAG 188
 QY 282 TACTTCTTTGACAGAAATTTTCAATTCGCTTGGCATTCATATTCCTTATAGC 341
 Db 187 TACTTCTTTGACAGAAATTTTCAATTCGCTTGGCATTCATATTCCTTATAGC 128
 QY 342 AGCCATGATTTCTTCTTCTTTTGGGAAATGTCCTAATGACATTTTCAATCTTTTGA 401
 Db 127 AGCCATGATTTCTTCTTCTTCTTTTGGGAAATGTCCTAATGACATTTTCAATCTTTTGA 68
 QY 402 TGTGACATTAATGCTTATTTGTTATGCTTATGCTTATGCTGA 441
 Db 67 TGTGACATTAATGCTTATTTGTTATGCTTATGCTTATGCTGA 28
 RESULT 6
 LOCUS BE767736/c 417 bp mRNA linear EST 20-SEP-2000
 DEFINITION RC3-GN0042-100800-011-e04 GN0042 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BE767736
 VERSION BE767736.1 GI:10221394
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 417)
 AUTHORS Dias Neto,E., Garcia Correa,R., Veijovskij-Almeida,S., Bionesi,M.R.,
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
 Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
 'M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
 Simpson,A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?cl=RC3-GN0042-100
 800-011-e04&t3=2000-08-10&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 37
 High quality sequence stop: 417.
 Location/Qualifiers
 1..417
 FEATURES
 source

TITLE
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
Simpson, A.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663

JOURNAL
MEDLINE
COMMENT
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?pl=PM3&t2=PM3-HT0909-
181000-010-905&t3=2000-10-18&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 39
High quality sequence stop: 403.
Location/Qualifiers
1. 403

FEATURES
Source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0909"
/dev_stage="Adult"
/note="Organ: head, neck; Vector: puc18; Site_1: Sma1;
Site_2: Sma1; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT
105 a 68 c 88 g 142 t
Origin

Query Match 27.8%; Score 369.4; DB 12; Length 403;
Best Local Similarity 99.7%; Pred. No. 1.9e-76;
Matches 370; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
68 TTACAAGCTCGAATTTTGCATGCGAATGGAATGTTAGGTTACTGGCTGGAATT 127
33 TTACAAGCTCGAATTTTGCATGCGAATGGAATGTTAGGTTACTGGCTGGAATT 92
128 TATCAGACTTGTGAGTAACAAGTTGAAGTTAGCAGATGAGGGGGAATATGAGCC 187
93 TATCAGACTTGTGAGTAACAAGTTGAAGTTAGCAGATGAGGGGGAATATGAGCC 152
188 TAAAGCTTAACAAATAATCAATCTGAGATAGTGGCTATGCTCCCAAGCCTTA 247
153 TAAAGCTTAACAAATAATCAATCTGAGATAGTGGCTATGCTCCCAAGCCTTA 212
248 TTGGGAACAGTTTTCCTGATGCTTGGAGAACTTTCTTTGACAGAAATTTTCAT 307
213 TTGGGAACAGTTTTCCTGATGCTTGGAGAACTTTCTTTGACAGAAATTTTCAT 272
308 TCTGCTTGCATTCATTTCTCCCTTTAGAGAGCATTGGATTTCTTTCTTTTGG 367
273 TCTGCTTGCATTCATTTCTCCCTTTAGAGAGCATTGGATTTCTTTCTTTTGG 332
368 GGAATGTCCCATTTAGATTTTTCAGATCTTTTATGTCATTAATGCAATTTGTAAT 427
333 GGAATGTCCCATTTAGATTTTTCAGATCTTTTATGTCATTAATGCAATTTGTAAT 392
428 GCCGTATTGG 438
393 GCCGTATTGG 403

RESULT 9
BG011354/0 408 bp MRNA linear EST 24-JAN-2001
LOCUS
DEFINITION
RC1-GN0268-091200-011-e10 GN0268 Homo sapiens cDNA, mRNA sequence.
ACCESSION
BG011354
VERSION
BG011354.1 GI:12459469
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

REFERENCE
AUTHORS
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
Simpson, A.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663

JOURNAL
MEDLINE
COMMENT
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?pl=RC1-GN0268-
091200-011-e10&t3=2000-12-09&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 10
High quality sequence stop: 408.
Location/Qualifiers
1. 408

FEATURES
Source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="GN0268"
/dev_stage="Adult"
/note="Organ: placenta, normal; Vector: puc18; Site_1: Sma1;
Site_2: Sma1; A mini-library was made by cloning
products derived from ORESTES PCR (U.S. Letters Patent
application No. 196,716 - Ludwig Institute for Cancer
Research) profiles into the puc 18 vector. Reverse
transcription of tissue mRNA and cDNA amplification were
performed under low stringency conditions."
BASE COUNT
142 a 90 c 71 g 105 t
Origin
Query Match 26.5%; Score 353; DB 12; Length 408;
Best Local Similarity 96.8%; Pred. No. 1.4e-72;
Matches 392; Conservative 0; Mismatches 10; Indels 3; Gaps 3;
42 TGTACCAACGTGAATTTTACAGTTTACAAATGTCTGGAATTTTGCATGCGATG 101
408 TGTACCAACGTGAATTTTACAGTTTACAAATGTCTGGAATTTTGCATGCGATG 349
102 AATGTTAAGTTTACTTGGCTGGAATTTATCAGATTTGAGTAACAAGTTGAATTTAG 161
348 AATGTTAAGTTTACTTGGCTGGAATTTATCAGATTTGAGTAACAAGTTGAATTTAG 289
162 CAGATGAGGGGGAATTTAGAGCCCTTAAGGCTTAACAATAATATCAGTATCTGAGATAG 221
288 CAGATGAGGGGGAATTTAGAGCCCTTAAGGCTTAACAATAATATCAGTATCTGAGATAG 229
222 TGGCTAATGTGGCTCCCGAGGCTTAATTTGGGAACAGTTTTCCTGATTTGAGAG 281
228 TGGCTAATGTGGCTCCCGAGGCTTAATTTGGGAACAGTTTTCCTGATTTGAGAG 169

QY 282 TACTTCTTTGACAGAAATTTTCATCTGCTGCATTCCTATATTCCTCTTTAG 341
 Db 168 TACTTCTTTGACAGAAATTTTCATCTGCTGCATTCCTATATTCCTCTTTAG 109
 QY 342 AGCATTTGATTTCTTTCTTTGTTGGAAATGCCATTAGCATTTGAGATCTTTGA 401
 Db 108 AGCCTGGGATTTCTTTCTTTGTTGGAAATGCCATTAGCATTTGAGATCTTTGA 50
 QY 402 TGTGCACTATGCAATATGTTGTAATGCCCTTTTGTGTAATCA 446
 Db 49 TGTGCACT-ATGCCATTATTTGGT-ATGCCGTTATTTGTGAGACACA 7
 RESULT 10
 BE929791 392 bp mRNA linear EST 02-OCT-2000
 LOCUS RC3-GN0042-280800-012-d09 GN0042 Homo sapiens cDNA, mRNA sequence.
 DEFINITION BE929791
 VERSION BE929791.1 GI:10455867
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 392)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, M. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
 Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-RC3-GN0042-280
 800-012-d09&tl3=2000-08-28&tl4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 23
 High quality sequence stop: 392.
 Location/Qualifiers
 1. 392
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="GN0042"
 /dev_stage="Adult"
 /note="Organ: placenta_normal; Vector: puc18; Site_1: SmaI
 ; Site_2: SmaI; A mini-library was made by cloning
 products derived from ORESTES PCR (U.S. letters Patent
 application No. 196,716 - Ludwig Institute for Cancer
 Research) profiles into the puc 18 vector. Reverse
 transcription of tissue mRNA and cDNA amplification were
 performed under low stringency conditions."
 BASE COUNT 105 a 68 c 89 g 129 t 1 others
 ORIGIN
 Query Match 26.5%; Score 352.8; DB 12; Length 392;
 Best Local Similarity 96.5%; Pred No. 1.5e-72;
 Matches 360; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
 QY 66 GTTACAAATGCTGGAATTTTGCACATGAGCAATGTTAGTACTTGGCTGGA 125
 Db 20 GGTTCACATGCTGGAATTTTGCACATGAGCAATGTTAGTACTTGGCTGGA 79

QY 126 TTATACAGCTTTGAGTAACAGACTTTAGCAGATGAGGGGAATATTGAGCC 185
 Db 80 TTATACAGCTTTGAGTAACAGACTTTAGCAGATGAGGGGAATATTGAGCC 139
 QY 186 CCTAAGGCTTAACAAATTAATCAGATCTGAGATGAGGCTAATGAGCTCCAGCC 245
 Db 140 CCTAAGGCTTAACAAATTAATCAGATCTGAGATGAGGCTAATGAGCTCCAGCC 199
 QY 246 AATTGGGACAGCTTTTCTGATGCTTTGAGAGACTTCTTTGACAGAAATTTTC 305
 Db 200 AATTGGGACAGCTTTTCTGATGCTTTGAGAGACTTCTTTGACAGAAATTTTC 259
 QY 306 ATCTGCTGCATTCGCTATATTCCTCTTTATAGAGCCATTGATTTCTTTG 365
 Db 260 ATCTGCTGCATTCGCTATATTCCTCTTTATAGAGCCATTGATTTCTTTG 319
 QY 366 TGGGAATGTCCTCCATTTAGCATTTTTCAGATCTTTGATGTCACATAATGCGTA 425
 Db 320 TGGGAATGTCCTCCATTTAGCATTTTTCAGATCTTTGATGTCACATAATGCGTA 379
 QY 426 ATGCCGTTATTGG 438
 Db 380 ATGCCGTTATTGG 392
 RESULT 11
 BG011351 402 bp mRNA linear EST 24-JAN-2001
 LOCUS R1-GN0268-091200-011-b01 GN0268 Homo sapiens cDNA, mRNA sequence.
 DEFINITION BG011351
 VERSION BG011351.1 GI:12459463
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 402)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, M. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
 Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=R1-GN0268-
 091200-011-b01&tl3=2000-12-09&tl4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 28
 High quality sequence stop: 402.
 Location/Qualifiers
 1. 402
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="GN0268"
 /dev_stage="Adult"
 /note="Organ: placenta_normal; Vector: puc18; Site_1: SmaI
 ; Site_2: SmaI; A mini-library was made by cloning
 products derived from ORESTES PCR (U.S. letters Patent
 application No. 196,716 - Ludwig Institute for Cancer

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: L1AM10228 Row: n Column: 16
 High quality sequence stop: 675.

FEATURES

Source

Location/Qualifiers
 1. 951

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4447695"
 /clone_id="NIH_MGC_89"
 /tissue_type="hypernephroma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: kidney; Vector: pCMV-Sport6; Site_1: Not; Site_2: SalI; cloned unidirectionally; oligo-dt primed. Average insert size 1.3 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC library."
 BASE COUNT 281 a 221 c 196 g 253 t
 ORIGIN

Query Match 25.2%; Score 335.2; DB 12; Length 951;
 Best Local Similarity 95.8%; Pred. No. 2.1e-68;
 Matches 366; Conservative 0; Mismatches 13; Indels 3; Gaps 2;

709 GACAGAGACTGTCTCAAAAAAGACATTATCATATTAATCTTATAGAGCCC 768
 76 GACAGAGACTGTCTCAAAAAAGACATTATCATATTAATCTTATAGAGCCC 135
 769 TATTTCTATCTGAAGGAGCTGTTTTTTTTTAAACAGTTAACTGATGTCACAG 828
 136 TATTTCTATCTGAAGGAGCTGTTTTTTTTTAAACAGTTAACTGATGTCACAG 193
 829 ACAAAATATTTCTGATCAGATAGTCCCTGTCAACAGTGAAGTGGTTTCATTAAGTG 888
 194 ACAAAATATTTCTGATCAGATAGTCCCTGTCAACAGTGAAGTGGTTTCATTAAGTG 253
 889 GGAAGAAAAAGCATTATTAAGTAACTTTTGGGAGAGTATTCAGTAATATTAAGT 948
 254 GGAAGAAAAAGCATTATTAAGTAACTTTTGGGAGAGTATTCAGTAATATTAAGT 312
 949 CTGGTCTCCCTTAAGAAAAAACCCCTTCACTTACTGTGATTTATATCCCTTAA 1008
 313 CTGGTCTCCCTTAAGAAAAAACCCCTTCACTTACTGTGATTTATATCCCTTAA 372
 1009 GTTCCAAAGTTAAATATCTTATTTCTGATATGCTTTATATCAAGAGCCTTATAGC 1068
 373 GTTCCAAAGTTAAATATCTTATTTCTGATATGCTTTATATCAAGAGCCTTATAGC 432
 1069 CAGTTCCGAGACACCACTATA 1090
 433 CTTGTACTACAGTATCTTTA 454

RESULT 14
 BF832034 396 bp mRNA linear EST 13-JAN-2001
 LOCUS PM3-HT0909-181000-011-h06 HT0909 Homo sapiens cDNA, mRNA sequence.
 DEFINITION BF832034
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 human.
 Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 396)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsushima, A., Bala, G.S., Simpson, D.H.,
 Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

JOURNAL
 MEDLINE
 COMMENT
 sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL:
<http://www.ludwig.org.br/scripts/gethtml2.pl?PL=PM3&ct=PM3-HT0909-181000-011-h06&rt=2000-10-18&ft=1>
 Seg primer: puc 18 forward
 High quality sequence stop: 87.
 Location/Qualifiers
 1. 396

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_id="HT0909"
 /dev_stage="Adult"
 /note="Organ: head, neck; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
 BASE COUNT 104 a 57 c 96 g 139 t
 ORIGIN

Query Match 25.0%; Score 333; DB 12; Length 396;
 Best Local Similarity 96.2%; Pred. No. 7e-68;
 Matches 364; Conservative 0; Mismatches 10; Indels 5; Gaps 4;

42 TGTACCAAGCGTAATTTACAGTGTTCACAAATGCTGGAATTTGGACGCGCATPAGG 101
 1 TGTACCAAGCGTAATTTACAGTGTTCACAAATGCTGGAATTTGGACGCGCATPAGG 60
 102 AATGTTAAGCTTACTTGGCTGGAATTTATCAGACT--TGTAGTAACAAAGTTGAAGTTT 159
 61 AATGTTAAGG-TACTTGGGTGGAATTTATGAAATGTGTAGTAACAAAGTTGAAGTTT 119
 160 AACAGATGAGGGGGAATTTATGAGGCCCTTAAGGCTAAACAAATATATCAGTATCGAGAT 219
 120 AACAGATGAGGGGGAATTTATGAGGCCCTTAAGGCTAAACAAATATATCAGTATCGAGAT 178
 220 AGTGGCTATGTGGCTCCCAAGGCCCTAATTTGGGAACAGTTTTCCTGATTTGTTGAGA 279
 179 AGTGGCTATGTGGCTCCCAAGGCCCTAATTTGGGAACAGTTTTCCTGATTTGTTGAGA 238
 280 AGTACTTCTTTTGACAGAAATTTTCATTCCTGCTGCTGATTTGCTATATTCCTCCCTTATA 339
 239 AGTACTTCTTTTGACAGAAATTTTCATTCCTGCTGCTGATTTGCTATATTCCTCCCTTATA 297
 340 GGAGCCATTTGATTTCTTTCCCTTTTGGGAATATGTCCTTATACATTTTCAGATCTTTT 399
 298 GGAGCCATTTGATTTCTTTCCCTTTTGGGAATATGTCCTTATACATTTTCAGATCTTTT 357
 400 GATGTCGACTAATGCCATTTATTTGTTATGTCGCTTATTTGG 438
 358 GATGTCGACTAATGCCATTTATTTGTTATGTCGCTTATTTGG 396

RESULT 15
 AA488304 376 bp mRNA linear EST 05-MAR-1998
 LOCUS AA488304
 DEFINITION
 5' similar to gb:M21121 T-CELL SPECIFIC RANTES PROTEIN PRECURSOR
 (HUMAN); mRNA sequence.
 ACCESSION AA488304

VERSION AA488304.1 GI:2215735
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 376)
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Gelsel, G., Jost, S.,
Kritzman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin
White, T., Wylie, T., Waterston, R., and Wilson, R.

TITLE WashU-NCI human EST Project
JOURNAL Unpublished (1997)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

FEATURES
Source This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert length: 1326 Std Error: 0.00
High quality sequence stop: 315.
Location/Qualifiers

1. 376

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:897294"

/clone_lib="Gessler Wilms tumor"

/sex="pooled (6)"

/lab_host="DH10b"

/note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; RNA
was prepared from a pool of 6 anonymous Wilms' tumor RNAs.
RNA was prepared by acid-phenol, followed by one round of
oligo dt selection. cDNA library preparation was with
the BRL/Life Tech. Superscript plasmid system. An
oligo-dT NotI primer for first strand synthesis generated
ggggcgccg(c) at the 3' end of the clones. A 5' SalI
adapter was used with sequence 5'-gtcagccagcgctcg-3'.
Resulting cDNAs were size selected (average size 2 kb).
NotI digested, and ligated into NotI/SalI-cut pSPORT1.
Library was constructed by Dr. Manfred Gessler."

BASE COUNT 97 a 77 c 75 g 127 t
ORIGIN

Query Match 24.8%; Score 330.4; DB 9; Length 376;
Best Local Similarity 95.5%; Pred. No. 2,9e-67;

Matches 340; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

OY 370 AATGTCCTATGACATTTTCAGATCTTTGATGTCACATAATGCCATTATGTAATGC 429
DB 1 AATGTCCTATGACATTTTCAGATCTTTGATGTCACATAATGCCATTATGTAATGC 60
OY 430 CCTATTGCGTAAATACAGATAGTTAAATTAACCTTACAGTAATCTTACACTGATTT 489
DB 61 CGTTATGCGTAAATACAGATAGTTAAATTAACCTTACAGTAATCTTACACTGATTT 120
OY 490 GCTGCACTCTACCAATAGCCTTTGAATGACGTAAAGTGAACAGAGAAAGAGCATG 549
DB 121 GCTGCACTCTACCAATAGCCTTTGAATGACGTAAAGTGAACAGAGAAAGAGCATG 180
OY 550 TCTGAGAAAGAGATAGTAAATATTTTGGTACTTTATCTGAATCCAAAGATCTGCTT 609
DB 181 TCTGAGAAAGAGATAGTAAATATTTTGGTACTTTATCTGAATCCAAAGATCTGCTT 240
OY 610 CCCCTGAGAGTTGTTTCTTCTTACGATCCTCATGTGAATCCCTCTGGGAGCAGACAG 669
DB 241 CCCCTGAGAGTTGTTTCTTCTTACGATCCTCATGTGAATCCCTCTGGGAGCAGAC 300
OY 670 AGTAGTAGAATCTCCATTTCTTCTTGTGTTTGAAGACAGAGACTGTCTC 725
DB 301 AGTAGTAGAATCTCCATTTCTTCTTGTGTTTGAAGACAGAGACTGTCTC 356

Search completed: June 27, 2003, 16:18:06
Job time: 1831 secs